



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
13.07.2022 Bulletin 2022/28

(51) International Patent Classification (IPC):
G06N 3/08 ^(2006.01) **G06N 3/04** ^(2006.01)

(21) Application number: **21203124.9**

(52) Cooperative Patent Classification (CPC):
G06N 3/0454; G06N 3/08

(22) Date of filing: **18.10.2021**

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
KH MA MD TN

(71) Applicant: **FUJITSU LIMITED**
Kawasaki-shi, Kanagawa 211-8588 (JP)

(72) Inventor: **YAMADA, Moyuru**
Kanagawa, 211-8588 (JP)

(74) Representative: **Hoffmann Eitle**
Patent- und Rechtsanwälte PartmbB
Arabellastraße 30
81925 München (DE)

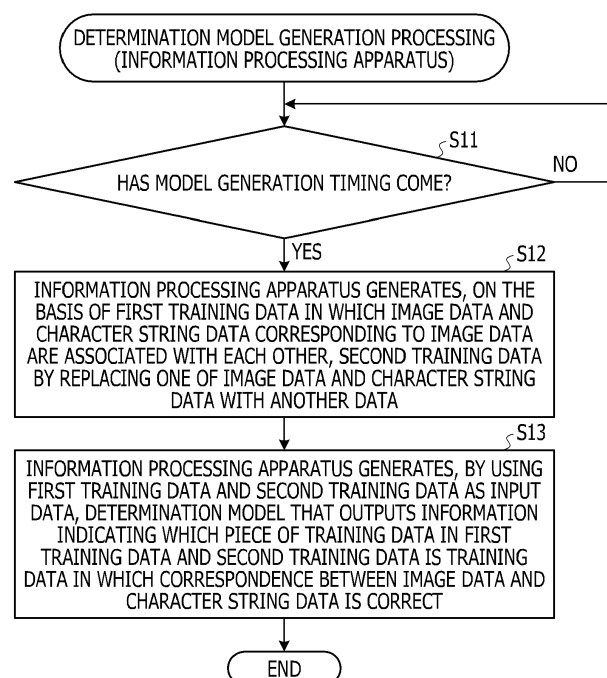
(30) Priority: **06.01.2021 JP 2021000937**

(54) **DETERMINATION MODEL GENERATION PROGRAM, INFORMATION PROCESSING APPARATUS, AND DETERMINATION MODEL GENERATION METHOD**

(57) A determination model generation program that causes at least one computer to execute a process, the process includes generating, based on first training data in which image data and character string data that corresponds to the image data are associated with each other, second training data by replacing one of data included in the first training data selected from the image

data and the character string data with another data; and generating, by using the first training data and the second training data as input data, a determination model that outputs information that indicates which training data selected from the first training data and the second training data is training data in which correspondence between the image data and the character string data is correct.

FIG. 8



Description

FIELD

5 **[0001]** The embodiment discussed herein is related to a determination model generation program, an information processing apparatus, and a determination model generation method.

BACKGROUND

10 **[0002]** In recent years, when an image that matches contents of a sentence (hereinafter, also referred to as a character string) or a sentence that matches contents of an image is searched for, a machine learning model (hereinafter, also referred to as a determination model) that determines a degree of matching between an image and a sentence may be used.

15 **[0003]** Such a determination model calculates a degree of matching between an image and a sentence by, for example, learning not only a pair of an image and sentence whose contents correspond to each other, but also a pair obtained by replacing one of the image and the sentence with another sample.

[0004] Jiasen Lu, Dhruv Batra, Devi Parikh, Stefan Lee, "ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks", URL: <https://arxiv.org/pdf/1908.02265.pdf> is disclosed as related art.

20 SUMMARY

[TECHNICAL PROBLEM]

25 **[0005]** Here, the determination model as described above is trained by inputting a plurality of pairs of sentences and images at the same time, for example. With this configuration, the determination model may proceed with learning about a relationship between pieces of data, as compared with a case where, for example, the determination model is trained by inputting a pair of a sentence and an image one by one.

30 **[0006]** However, for example, in a case where learning is performed by using a plurality of pairs in which both images and sentences are randomly determined, the determination model may not be able to sufficiently learn about characteristics of each pair, and may not be able to calculate a degree of matching between an image and a sentence with sufficient accuracy.

35 **[0007]** Therefore, in one aspect, an embodiment aims to provide a determination model generation program, an information processing apparatus, and a determination model generation method that enable generation of a determination model that accurately determines correspondence between an image and a sentence.

[SOLUTION TO PROBLEM]

40 **[0008]** According to an aspect of the embodiments, a determination model generation program that causes at least one computer to execute a process, the process includes generating, based on first training data in which image data and character string data that corresponds to the image data are associated with each other, second training data by replacing one of data included in the first training data selected from the image data and the character string data with another data; and generating, by using the first training data and the second training data as input data, a determination model that outputs information that indicates which training data selected from the first training data and the second training data is training data in which correspondence between the image data and the character string data is correct.

45 [ADVANTAGEOUS EFFECTS OF INVENTION]

50 **[0009]** According to one aspect, it is possible to generate a determination model that accurately determines correspondence between an image and a sentence.

BRIEF DESCRIPTION OF DRAWINGS

[0010]

55 FIG. 1 is a diagram illustrating a configuration of an information processing system 10;
 FIG. 2 is a diagram illustrating a determination model MD11 in a first comparative example;
 FIG. 3 is a diagram illustrating the determination model MD11 in the first comparative example;
 FIG. 4 is a diagram illustrating a determination model MD12 in a second comparative example;

FIG. 5 is a diagram illustrating the determination model MD12 in the second comparative example;
 FIG. 6 is a diagram illustrating a hardware configuration of an information processing apparatus 1;
 FIG. 7 is a block diagram of functions of the information processing apparatus 1;
 FIG. 8 is a flowchart illustrating an outline of determination model generation processing in a first embodiment;
 FIG. 9 is a diagram illustrating processing of a determination model MD1 in the first embodiment;
 FIG. 10 is a diagram illustrating processing of the determination model MD1 in the first embodiment;
 FIG. 11 is a flowchart illustrating details of the determination model generation processing in the first embodiment;
 FIG. 12 is a flowchart illustrating the details of the determination model generation processing in the first embodiment;
 FIG. 13 is a flowchart illustrating the details of the determination model generation processing in the first embodiment;
 FIG. 14 is a flowchart illustrating the details of the determination model generation processing in the first embodiment;
 FIG. 15 is a flowchart illustrating the details of the determination model generation processing in the first embodiment;
 FIG. 16 is a flowchart illustrating the details of the determination model generation processing in the first embodiment;
 FIG. 17 is a flowchart illustrating the details of the determination model generation processing in the first embodiment;
 FIG. 18 is a flowchart illustrating the details of the determination model generation processing in the first embodiment;
 FIG. 19 is a diagram illustrating processing of a determination model MD2 in the first embodiment; and
 FIG. 20 is a diagram illustrating the processing of the determination model MD2 in the first embodiment.

DESCRIPTION OF EMBODIMENTS

[Configuration of Information Processing System]

[0011] First, a configuration of an information processing system 10 will be described. FIG. 1 is a diagram illustrating the configuration of the information processing system 10.

[0012] The information processing system 10 illustrated in FIG. 1 includes, for example, an information processing apparatus 1 including one or more physical machines, and an operation terminal 2 in which an operator who generates a determination model (hereinafter, also simply referred to as an operator) performs input of necessary information, or the like. The operation terminal 2 may be, for example, a personal computer (PC). Furthermore, the information processing apparatus 1 and the operation terminal 2 are connected via a network NW such as the Internet, for example.

[0013] The information processing apparatus 1 generates a determination model by using, for example, a plurality of pieces of training data prepared in advance. Hereinafter, determination models in comparative examples will be described.

[Determination Model in First Comparative Example]

[0014] FIGs. 2 and 3 are diagrams illustrating a determination model MD11 in a first comparative example. For example, FIG. 2 is a diagram illustrating processing of the determination model MD11 in a learning stage. Furthermore, FIG. 3 is a diagram illustrating processing of the determination model MD11 in an inference stage. A first neural network NN1 may be, for example, Transformer. Furthermore, a second neural network NN2 may be, for example, a liner network for learning a score function.

[0015] Note that, hereinafter, description will be made assuming that each determination model includes the neural network NN1 (hereinafter, also referred to as the first neural network NN1) and the neural network NN2 (hereinafter, also referred to as the second neural network NN2).

[0016] First, the processing of the determination model MD11 in the learning stage will be described.

[0017] As illustrated in FIG. 2, the information processing apparatus 1 inputs, to the first neural network NN1, a pair of image data IM1 and character string data ST1 that are included in training data for learning. For example, the information processing apparatus 1 inputs, to the first neural network NN1, a pair of image data IM1 and character string data ST1 whose contents correspond to each other, or a pair of image data IM1 and character string data ST1 whose contents do not correspond.

[0018] Then, in response to reception of input of the pair of image data IM1 and character string data ST1, the first neural network NN1 calculates and outputs a vector indicating characteristics of the image data IM1 (hereinafter, also referred to as an image vector) and a vector indicating characteristics of the character string data ST1 (hereinafter, also referred to as a character string vector).

[0019] Subsequently, for example, the information processing apparatus 1 calculates an element product of the image vector and the character string vector that are output from the first neural network NN1, and inputs the element product to the second neural network NN2.

[0020] Then, for example, in response to reception of input of the element product, the second neural network NN2 calculates and outputs a degree of matching between the contents indicated by the image data IM1 and the contents indicated by the character string data ST1.

[0021] Thereafter, for example, the information processing apparatus 1 adjusts weights of the first neural network NN1

and the second neural network NN2 such that an error between the degree of matching output by the second neural network and a value (correct data) indicating whether or not the contents indicated by the image data IM1 and the contents indicated by the character string data ST1 correspond to each other becomes small.

[0022] Next, the processing of the determination model MD11 in the inference stage will be described.

[0023] As illustrated in FIG. 3, for example, in a case where training data to be inferred (hereinafter, also referred to as new training data) is input via the operation terminal 2, the information processing apparatus 1 inputs a pair of image data IM2 and character string data ST2 that are included in the new training data to the first neural network NN1.

[0024] Then, in response to reception of input of the pair of image data IM2 and character string data ST2, the first neural network NN1 calculates and outputs an image vector indicating characteristics of the image data IM2 and a character string vector indicating characteristics of the character string data ST2.

[0025] Subsequently, for example, the information processing apparatus 1 calculates an element product of the image vector and the character string vector that are output from the first neural network NN1, and inputs the element product to the second neural network NN2.

[0026] Then, for example, in response to reception of input of the element product, the second neural network NN2 calculates and outputs a degree of matching between contents indicated by the image data IM2 and contents indicated by the character string data ST2.

[0027] Thereafter, the information processing apparatus 1 outputs, for example, the degree of matching output by the second neural network to the operation terminal 2 as a degree of matching between the image data IM2 and the character string data ST2 that are included in the new training data.

[0028] Here, for example, in a case where training is performed by inputting a pair of image data IM1 and character string data ST1 one by one, it is not possible for the determination model MD11 to totally optimize learning, for example, it is not possible to learn about a relationship between pieces of data. Therefore, for example, in a case where image data IM2 and character string data ST2 related to the same object are input in the inference stage, the determination model MD11 may output a high degree of matching even in a case where contents of the image data IM2 and contents of the character string data ST2 indicate different situations.

[Determination Model in Second Comparative Example]

[0029] Next, a determination model MD12 in a second comparative example will be described. FIGs. 4 and 5 are diagrams illustrating the determination model MD12 in the second comparative example. For example, FIG. 4 is a diagram illustrating processing of the determination model MD12 in a learning stage. Furthermore, FIG. 5 is a diagram illustrating processing of the determination model MD12 in an inference stage.

[0030] Note that, hereinafter, image data IM1a, image data IM1b, and image data IM1c are also collectively referred to simply as image data IM1, and character string data ST1a, character string data ST1b, and character string data ST1c are also collectively referred to simply as character string data ST1.

[0031] First, the processing of the determination model MD12 in the learning stage will be described.

[0032] As illustrated in FIG. 4, the information processing apparatus 1 inputs, to the first neural network NN1, a plurality of pairs of image data IM1 and character string data ST1 that are included in training data for learning. For example, the information processing apparatus 1 inputs, to the first neural network NN1, one pair of image data IM1 and character string data ST1 whose contents correspond to each other, and a plurality of pairs of image data IM1 and character string data ST1 whose contents do not correspond.

[0033] For example, as illustrated in FIG. 4, the information processing apparatus 1 inputs, to the first neural network NN1, a plurality of pairs including a pair of image data IM1a and character string data ST1a whose contents correspond to each other, a pair of image data IM1b and character string data ST1b whose contents do not correspond, and a pair of image data IM1c and character string data ST1c whose contents do not correspond.

[0034] Then, in response to reception of input of the pairs of image data IM1 and character string data ST1, the first neural network NN1 calculates and outputs image vectors indicating characteristics of the image data IM1 and character string vectors indicating characteristics of the character string data ST1.

[0035] Subsequently, for example, the information processing apparatus 1 calculates element products of the image vectors and the character string vectors that are output from the first neural network NN1, and inputs the element products to the second neural network NN2.

[0036] Then, for example, in a case where input of the plurality of element products is received, the second neural network NN2 calculates and outputs, for each pair of image data IM1 and character string data ST1, a degree of matching between the image data IM1 and the character string data ST1 that are included in each pair.

[0037] Thereafter, for example, the information processing apparatus 1 adjusts the weights of the first neural network NN1 and the second neural network NN2 such that a classification error (cross entropy) calculated from the plurality of degrees of matching output by the second neural network and information indicating the pair of image data IM1 and character string data ST1 whose contents correspond to each other (hereinafter, also referred to as a correct pair)

becomes small.

[0038] Next, the processing of the determination model MD12 in the inference stage will be described.

[0039] As illustrated in FIG. 5, for example, in a case where new training data is input via the operation terminal 2, the information processing apparatus 1 inputs, to the first neural network NN1, a pair of image data IM2 and character string data ST2 that are included in the new training data.

[0040] Then, in response to reception of input of the pair of image data IM2 and character string data ST2, the first neural network NN1 calculates and outputs an image vector indicating characteristics of the image data IM2 and a character string vector indicating characteristics of the character string data ST2.

[0041] Subsequently, for example, the information processing apparatus 1 calculates an element product of the image vector and the character string vector that are output from the first neural network NN1, and inputs the element product to the second neural network NN2.

[0042] Then, for example, in response to reception of input of the element product, the second neural network NN2 calculates and outputs a degree of matching between contents indicated by the image data IM2 and contents indicated by the character string data ST2.

[0043] Thereafter, the information processing apparatus 1 outputs, for example, the degree of matching output by the second neural network to the operation terminal 2 as a degree of matching between the image data IM2 and the character string data ST2 that are included in the new training data.

[0044] For example, the determination model MD12 is trained by using the plurality of pairs of image data IM1 and character string data ST1 at the same time. With this configuration, the determination model MD12 may proceed with learning of a relationship between pieces of training data as compared with the determination model MD11 described with reference to FIGs. 2 and 3, and may perform learning that is totally optimized.

[0045] However, for example, in a case where training is performed by using a plurality of pairs in which both the image data IM1 and the character string data ST1 are randomly determined, it is not possible for the determination model MD12 to sufficiently learn about characteristics of each pair. Thus, it may not possible for the determination model MD12 to calculate the degree of matching between the image data IM2 and the character string data ST2 with sufficient accuracy in the inference stage.

[0046] Therefore, the information processing apparatus 1 in the present embodiment generates, on the basis of training data (hereinafter, also referred to as first training data) in which image data IM1 and character string data ST1 corresponding to the image data IM1 are associated with each other, training data (hereinafter, also referred to as second training data) by replacing one of the image data IM1 and the character string data ST1 that are included in the first training data with another data.

[0047] Then, the information processing apparatus 1 generates, by using the first training data and the second training data as input data, a determination model that outputs information indicating which piece of training data in the first training data and the second training data is training data in which correspondence between the image data IM1 and the character string data ST1 is correct.

[0048] For example, the information processing apparatus 1 in the present embodiment generates, in a case where a plurality of pairs of image data IM1 and character string data ST1 to be input to the determination model is generated, the plurality of pairs by combining different character string data ST1 with the same image data IM1. Furthermore, in this case, the information processing apparatus 1 generates the plurality of pairs by combining different image data IM1 with the same character string data ST1. Then, the information processing apparatus 1 generates the determination model by causing the determination model to learn a degree of similarity calculated for each of the generated plurality of pairs.

[0049] With this configuration, the information processing apparatus 1 may cause a determination model to learn a more detailed relationship between pieces of training data, and may generate a determination model that accurately determines correspondence between image data IM1 and character string data ST1.

[Hardware Configuration of Information Processing Apparatus]

[0050] Next, a hardware configuration of the information processing apparatus 1 will be described. FIG. 6 is a diagram illustrating the hardware configuration of the information processing apparatus 1.

[0051] As illustrated in FIG. 6, the information processing apparatus 1 includes a central processing unit (CPU) 101 as a processor, a memory 102, a communication device 103, and a storage medium 104. Each of the units is connected to each other via a bus 105.

[0052] The storage medium 104 includes, for example, a program storage area (not illustrated) that stores a program 110 for performing determination model generation processing. Furthermore, the storage medium 104 includes, for example, an information storage area 130 that stores information used when the determination model generation processing is performed. Note that the storage medium 104 may be, for example, a hard disk drive (HDD) or a solid state drive (SSD).

[0053] The CPU 101 executes the program 110 loaded from the storage medium 104 into the memory 102 to perform the determination model generation processing.

[0054] Furthermore, the communication device 103 communicates with the operation terminal 2 via the network NW, for example.

[Functions of Information Processing Apparatus]

[0055] Next, functions of the information processing apparatus 1 will be described. FIG. 7 is a block diagram of the functions of the information processing apparatus 1.

[0056] As illustrated in FIG. 7, the information processing apparatus 1 implements various functions including a data reception unit 111, a data management unit 112, a data generation unit 113, a vector generation unit 114, a similarity calculation unit 115, a model training unit 116, a matching degree calculation unit 117, and a result output unit 118, for example, by hardware such as the CPU 101 and the memory 102 organically cooperating with the program 110.

[0057] Furthermore, as illustrated in FIG. 7, the information processing apparatus 1 stores, for example, first training data DT1, second training data DT2, and new training data DT3 in the information storage area 130.

First, the functions in a learning stage will be described.

[0058] The data reception unit 111 receives, for example, the first training data DT1 transmitted from the operation terminal 2. The first training data DT1 is training data including image data IM1 and character string data ST1 whose contents correspond to each other. Then, the data management unit 112 stores, for example, the first training data DT1 received by the data reception unit 111 in the information storage area 130.

[0059] The data generation unit 113 generates the second training data DT2 on the basis of the first training data DT1 stored in the information storage area 130. The second training data DT2 is training data obtained by replacing one of the image data IM1 and the character string data ST1 that are included in the first training data DT1 with another data. For example, the second training data DT2 is training data including image data IM1 and character string data ST1 whose contents do not correspond. Then, the data management unit 112 stores, for example, the second training data DT2 generated by the data generation unit 113 in the information storage area 130.

[0060] The vector generation unit 114 generates an image vector indicating characteristics of each of the image data IM1 included in the first training data DT1 and the second training data DT2 by inputting the first training data DT1 and the second training data DT2 that are stored in the information storage area 130 to the first neural network NN1. Furthermore, the vector generation unit 114 generates character string vectors indicating characteristics of the character string data ST1 included in the first training data DT1 and the second training data DT2 by inputting the first training data DT1 and the second training data DT2 that are stored in the information storage area 130 to the first neural network NN1.

[0061] The similarity calculation unit 115 calculates, for example, an inner product of an image vector and a character string vector that are generated by the vector generation unit 114 from the first training data DT1 as a degree of similarity between the image data IM1 and the character string data ST1 that are included in the first training data DT1. Furthermore, the similarity calculation unit 115 calculates, for example, an inner product of an image vector and a character string vector that are generated by the vector generation unit 114 from the second training data DT2 as a degree of similarity between the image data IM1 and the character string data ST1 that are included in the second training data DT2.

[0062] The model training unit 116 calculates, for example, a classification error from each degree of similarity calculated by the similarity calculation unit 115 and information indicating the first training data DT1 (hereinafter, also referred to as first information). Then, the model training unit 116 adjusts (learns) the weight of the first neural network NN1 such that the calculated classification error becomes small.

[0063] The matching degree calculation unit 117 calculates, for example, an element product of an image vector and a character string vector that are generated by the vector generation unit 114 from the first training data DT1. Furthermore, the matching degree calculation unit 117 calculates, for example, an element product of an image vector and a character string vector that are generated by the vector generation unit 114 from the second training data DT2. For example, the matching degree calculation unit 117 calculates an element product for one piece of the first training data DT1 and element products for a plurality of pieces of second training data DT2.

[0064] Then, the matching degree calculation unit 117 calculates, for example, a degree of matching between the image data IM1 and the character string data ST1 that are included in the first training data DT1 by inputting the element product for the first training data DT1 to the second neural network NN2. Furthermore, the matching degree calculation unit 117 calculates, for example, degrees of matching between the image data IM1 and the character string data ST1 that are included in the second training data DT2 by inputting the element products for the second training data DT2 to the second neural network NN2.

[0065] Moreover, for example, the model training unit 116 adjusts the weights of the first neural network NN1 and the second neural network NN2 such that an error between the degree of matching for the first training data DT1 and a

value corresponding to the first training data DT1 (hereinafter, also referred to as second information) becomes small. Furthermore, for example, the model training unit 116 adjusts the weights of the first neural network NN1 and the second neural network NN2 such that an error between the degrees of matching for the second training data DT2 and a value corresponding to the second training data DT2 becomes small.

[0066] Next, the functions in an inference stage will be described.

[0067] The data reception unit 111 receives, for example, the new training data DT3 transmitted from the operation terminal 2. Then, the data management unit 112 stores, for example, the new training data DT3 received by the data reception unit 111 in the information storage area 130.

[0068] The vector generation unit 114 generates an image vector indicating characteristics of image data IM2 included in the new training data DT3 by inputting the new training data DT3 received by the data reception unit 111 to the first neural network NN1. Furthermore, the vector generation unit 114 generates a character string vector indicating characteristics of character string data ST2 included in the new training data DT3 by inputting the new training data DT3 to the first neural network NN1.

[0069] The similarity calculation unit 115 calculates, for example, an inner product of an image vector and a character string vector that are generated by the vector generation unit 114 from the new training data DT3 as a degree of similarity between the image data IM2 and the character string data ST2 that are included in the new training data DT3.

[0070] The matching degree calculation unit 117 calculates, for example, an element product of an image vector and a character string vector that are generated by the vector generation unit 114 from the new training data DT3.

[0071] Then, the matching degree calculation unit 117 calculates, for example, a degree of matching between the image data IM2 and the character string data ST2 that are included in the new training data DT3 by inputting the element product for the new training data DT3 to the second neural network NN2.

[0072] The result output unit 118 outputs, for example, a degree of similarity calculated by the similarity calculation unit 115 to the operation terminal 2 as a degree of similarity between the image data IM2 and the character string data ST2 that are included in the new training data DT3. Furthermore, the result output unit 118 outputs, for example, a degree of matching calculated by the matching degree calculation unit 117 to the operation terminal 2 as a degree of matching between the image data IM2 and the character string data ST2 that are included in the new training data DT3.

[0073] Note that, hereinafter, a case where the matching degree calculation unit 117 calculates an element product of an image vector and a character string vector will be described. However, the matching degree calculation unit 117 may calculate, for example, the vector sum of an image vector and a character string vector. Then, in this case, the matching degree calculation unit 117 may input the calculated vector sum to the second neural network NN2.

[Outline of First Embodiment]

[0074] Next, an outline of a first embodiment will be described. FIG. 8 is a flowchart illustrating an outline of determination model generation processing in the first embodiment.

[0075] As illustrated in FIG. 8, the information processing apparatus 1 stands by until, for example, a model generation timing comes (NO in S11). The model generation timing may be, for example, a timing at which an operator inputs that a determination model is to be generated.

[0076] Then, in a case where the model generation timing comes (YES in S11), the information processing apparatus 1 generates, on the basis of the first training data DT1 in which the image data IM1 and the character string data ST1 corresponding to the image data IM1 are associated with each other, the second training data DT2 by replacing one of the image data IM1 and the character string data ST1 with another data (S12).

[0077] Moreover, the information processing apparatus 1 generates, by using the first training data DT1 and the second training data DT2 as input data, a determination model that outputs information indicating which piece of training data in the first training data DT1 and the second training data DT2 is training data in which correspondence between the image data IM1 and the character string data ST1 is correct (S13).

[0078] With this configuration, the information processing apparatus 1 may cause the determination model to learn a detailed relationship between pieces of the training data. Therefore, the information processing apparatus 1 may generate the determination model that accurately determines correspondence between the image data IM1 and the character string data ST1.

[0079] Furthermore, by replacing one of the image data IM1 and the character string data ST1 that are included in the first training data DT1 with another data to generate the second training data DT2, the information processing apparatus 1 may suppress increase in a work load and work time needed to generate the second training data DT2.

[Specific Example (1) of Determination Model]

[0080] Next, a determination model MD1 in the first embodiment will be described. FIGs. 9 and 10 are diagrams illustrating processing of the determination model MD1 in the first embodiment.

[0081] First, processing of the determination model MD1 in a learning stage will be described.

[0082] As illustrated in FIG. 9, the information processing apparatus 1 inputs, to the first neural network NN1, a pair of image data IM1 and character string data ST1 that are included in the first training data DT1 and pairs of image data IM1 and character string data ST1 that are included in the second training data DT2.

[0083] For example, as illustrated in FIG. 9, the information processing apparatus 1 inputs, to the first neural network NN1, a plurality of pairs including a pair of image data IM1a and character string data ST1a that are included in the first training data DT1, a pair of image data IM1b and character string data ST1b that are included in the second training data DT2, and a pair of image data IM1c and character string data ST1c that are included in the second training data DT2.

[0084] Then, for example, in response to reception of input of the pair of image data IM1a and character string data ST1a, the first neural network NN1 calculates and outputs an image vector indicating characteristics of the image data IM1a and a character string vector indicating characteristics of the character string data ST1a. Similarly, for example, in response to reception of input of the pair of image data IM1b and character string data ST1b, the first neural network NN1 calculates and outputs an image vector indicating characteristics of the image data IM1b and a character string vector indicating characteristics of the character string data ST1b. Moreover, for example, in response to reception of input of the pair of image data IM1c and character string data ST1c, the first neural network NN1 calculates and outputs an image vector indicating characteristics of the image data IM1c and a character string vector indicating characteristics of the character string data ST1c.

[0085] Next, the information processing apparatus 1 calculates, for example, an inner product of the image vector corresponding to the image data IM1a and the character string vector corresponding to the character string data ST1a as a degree of similarity between the image data IM1a and the character string data ST1a. Similarly, the information processing apparatus 1 calculates, for example, an inner product of the image vector corresponding to the image data IM1b and the character string vector corresponding to the character string data ST1b as a degree of similarity between the image data IM1b and the character string data ST1b. Moreover, the information processing apparatus 1 calculates, for example, an inner product of the image vector corresponding to the image data IM1c and the character string vector corresponding to the character string data ST1c as a degree of similarity between the image data IM1c and the character string data ST1c.

[0086] Thereafter, for example, the information processing apparatus 1 adjusts the weight of the first neural network NN1 such that a classification error calculated from the degree of similarity for each of the first training data DT1 and the second training data DT2 and information specifying the pair (correct pair) of image data IM1 and character string data ST1 that are included in the first training data DT1 becomes small.

[0087] Next, processing of the determination model MD1 in an inference stage will be described.

[0088] As illustrated in FIG. 10, the information processing apparatus 1 inputs, to the first neural network NN1, a pair of image data IM2 and character string data ST2 that are included in the new training data DT3.

[0089] Then, for example, in response to reception of input of the pair of image data IM2 and character string data ST2, the first neural network NN1 calculates and outputs an image vector indicating characteristics of the image data IM2 and a character string vector indicating characteristics of the character string data ST2.

[0090] Thereafter, the information processing apparatus 1 calculates, for example, an inner product of the image vector corresponding to the image data IM2 and the character string vector corresponding to the character string data ST2 as a degree of similarity between the image data IM2 and the character string data ST2.

[0091] Then, the information processing apparatus 1 outputs, for example, the calculated degree of similarity to the operation terminal 2 as a degree of similarity between the image data IM2 and the character string data ST2 that are included in the new training data DT3.

[0092] For example, unlike the case described with reference to FIGs. 2 to 5, the information processing apparatus 1 generates the determination model MD1 using only the first neural network NN1 (determination model MD1 having no second neural network NN2).

[0093] With this configuration, the information processing apparatus 1 may generate the first neural network NN1 capable of outputting an image vector and character string vector different for each piece of training data.

[Details of First Embodiment]

[0094] Next, details of the first embodiment will be described. FIGs. 11 to 18 are flowcharts illustrating details of the determination model generation processing in the first embodiment. Furthermore, FIGs. 19 and 20 are diagrams illustrating the details of the determination model generation processing in the first embodiment.

[Data Management Processing]

[0095] First, in the determination model generation processing, processing of storing the first training data DT1 in the information storage area 130 (hereinafter, also referred to as data management processing) will be described. FIG. 11

is a flowchart illustrating the data management processing.

[0096] As illustrated in FIG. 11, for example, the data reception unit 111 stands by until receiving a plurality of pieces of training data DT1 transmitted from the operation terminal 2 (NO in S21).

[0097] Then, in the case of receiving the plurality of pieces of training data DT1 (YES in S21), the data management unit 112 stores the plurality of pieces of training data DT1 received in the processing in S21 in the information storage area 130 (S22).

[Main Processing of Determination Model Generation Processing]

[0098] Next, main processing of the determination model generation processing will be described. FIGs. 12 to 16 are flowcharts illustrating the main processing of the determination model generation processing.

[0099] As illustrated in FIG. 12, the data generation unit 113 stands by until a model generation timing comes (NO in S31).

[0100] Then, in a case where the model generation timing comes (YES in S31), the data generation unit 113 specifies any one piece of first training data DT1 among the plurality of pieces of first training data DT1 stored in the information storage area 130 (S32).

[0101] Subsequently, the data generation unit 113 generates R pieces of second training data DT2 by replacing one of image data IM1 and character string data ST1 that are included in the first training data DT1 specified in the processing in S32 with another data (S33). Hereinafter, details of the processing in S33 will be described.

[Details (1) of Processing in S33]

[0102] First, a first example of the processing in S33 will be described. FIG. 15 is a flowchart illustrating the details of the processing in S33.

[0103] As illustrated in FIG. 15, for example, the data generation unit 113 duplicates R pieces of image data IM1 included in the first training data DT1 specified in the processing in S32 (S61).

[0104] Then, for example, from a plurality of pieces of character string data ST1 included in the plurality of pieces of first training data DT1 stored in the information storage area 130, the data generation unit 113 specifies R pieces of character string data ST1 other than character string data ST1 included in the first training data DT1 specified in the processing in S32 (S62).

[0105] Thereafter, for example, the data generation unit 113 generates R pieces of second training data DT2 by associating the R pieces of image data IM1 duplicated in the processing in S61 with the R pieces of character string data ST1 specified in the processing in S62 (S63).

[Details (2) of Processing in S33]

[0106] Next, a second example of the processing in S33 will be described. FIG. 16 is another flowchart illustrating the details of the processing in S33.

[0107] For example, the data generation unit 113 duplicates R pieces of character string data ST1 included in the first training data DT1 specified in the processing in S32 (S71).

[0108] Then, for example, from a plurality of pieces of image data IM1 included in the plurality of pieces of first training data DT1 stored in the information storage area 130, the data generation unit 113 specifies R pieces of image data IM1 other than image data IM1 included in the first training data DT1 specified in the processing in S32 (S72).

[0109] Thereafter, for example, the data generation unit 113 generates R pieces of second training data DT2 by associating the R pieces of character string data ST1 duplicated in the processing in S71 with the R pieces of image data IM1 specified in the processing in S72 (S73).

[0110] For example, the information processing apparatus 1 may facilitate learning about a relationship between pieces of training data by generating a determination model by using a plurality of pieces of second training data DT2 in which either image data IM1 or character string data ST1 is matched.

[0111] Therefore, the information processing apparatus 1 may generate the determination model that accurately determines correspondence between the image data IM1 and the character string data ST1.

[0112] Note that, for example, the data generation unit 113 may generate a part of the R pieces of second training data DT2 by performing the processing in S61 to S63, and may generate another part of the R pieces of second training data DT2 by performing the processing in S71 to S73.

[0113] Furthermore, for example, the data generation unit 113 may generate the R pieces of second training data DT2 by performing the processing in S61 to S63 and the processing in S71 to S73 in parallel.

[0114] Returning to FIG. 12, the vector generation unit 114 calculates, by using the first neural network NN1, an image vector for the image data IM1 and a character string vector for the character string data ST1 for each of the first training

data DT1 specified in the processing in S32 and the R pieces of second training data DT2 generated in the processing in S33 (S34).

[0115] For example, the vector generation unit 114 inputs, to the first neural network NN1, each of the first training data DT1 specified in the processing in S32 and the R pieces of second training data DT2 generated in the processing in S33. Then, the vector generation unit 114 acquires each of the image vectors and the character string vectors that are output from the first neural network NN1 as an image vector and a character string vector for each of the first training data DT1 specified in the processing in S32 and the R pieces of second training data DT2 generated in the processing in S33.

[0116] Then, for each of the first training data DT1 specified in the processing in S32 and the R pieces of second training data DT2 generated in the processing in S33, the similarity calculation unit 115 calculates a degree of similarity between the image data IM1 and the character string data ST1 by calculating an inner product of the vectors calculated in the processing in S34 (S35).

[0117] Subsequently, as illustrated in FIG. 13, the model training unit 116 adjusts the weight of the first neural network NN1 such that a classification error calculated from each degree of similarity calculated in the processing in S35 becomes small (S41).

[0118] For example, the model training unit 116 adjusts the weight of the first neural network NN1 such that a classification error L indicated in the following Equation (1) becomes small.

[Expression 1]

[0119]

$$L = -\log \frac{e^{E_p}}{\sum_{i=0}^R e^{E_i}} \quad \cdot \cdot \cdot \quad (1)$$

[0120] In the above Equation (1), E_i indicates a degree of similarity corresponding to an i-th piece of training data among the first training data DT1 specified in the processing in S32 and the R pieces of second training data DT2 generated in the processing in S33. Furthermore, E_p indicates a degree of similarity corresponding to the first training data DT1 specified in the processing in S32.

[0121] Next, the matching degree calculation unit 117 specifies any one piece of second training data DT2 included in the R pieces of second training data DT2 generated in the processing in S33 (S42).

[0122] Then, for each of the first training data DT1 specified in the processing in S32 and the second training data DT2 specified in the processing in S42, the matching degree calculation unit 117 calculates an element product of the vectors calculated in the processing in S34 (S43).

[0123] Moreover, the matching degree calculation unit 117 calculates, by using the second neural network NN2, a degree of matching between the image data IM1 and the character string data ST1 for each of an element product corresponding to the first training data DT1 specified in the processing in S32 and an element product corresponding to the second training data DT2 specified in the processing in S42 (S44).

[0124] For example, the matching degree calculation unit 117 inputs each of the element products calculated in the processing in S44 to the second neural network NN2. Then, the matching degree calculation unit 117 acquires the degrees of matching output from the second neural network NN2 as a degree of matching between the image data IM1 and the character string data ST1 that are included in the first training data DT1 specified in the processing in S32 and a degree of matching between the image data IM1 and the character string data ST1 that are included in the second training data DT2 specified in the processing in S42.

[0125] Subsequently, as illustrated in FIG. 14, the model training unit 116 calculates an error between the degree of matching for the first training data DT1 specified in the processing in S32 and a value corresponding to the first training data DT1 (S51).

[0126] Furthermore, the model training unit 116 calculates an error between the degree of matching for the second training data DT2 specified in the processing in S42 and a value corresponding to the second training data DT2 (S52).

[0127] Thereafter, the model training unit 116 adjusts the weights of the first neural network NN1 and the second neural network NN2 such that the errors calculated in the processing in S51 and S52 become small (S53).

[0128] For example, in a case where the degree of matching for the first training data DT1 specified in the processing in S32 is a value between 0 and 1, and the value corresponding to the first training data DT1 is 1, the model training unit 116 adjusts the weight of the second neural network NN2 such that the degree of matching for the first training data DT1 specified in the processing in S32 approaches 1. Furthermore, the model training unit 116 adjusts the weights of

the first neural network NN1 and the second neural network NN2 such that the error between the degree of matching for the second training data DT2 specified in the processing in S42 and the value corresponding to the second training data DT2 becomes small.

[0129] Then, the model training unit 116 determines whether or not the first neural network NN1 and the second neural network NN2 satisfy a predetermined condition (S54).

[0130] For example, the model training unit 116 calculates the sum of the error calculated in the processing in S51 (error for the first training data DT1) and the error calculated in the processing in S52 (error for the second training data DT2). Subsequently, the model training unit 116 calculates, for example, an average of the calculated sum and the classification error calculated from each degree of similarity calculated in the processing in S35. Then, for example, in a case where the calculated average is below a predetermined threshold, the model training unit 116 determines that the first neural network NN1 and the second neural network NN2 satisfy the predetermined condition.

[0131] As a result, in a case where it is determined that the first neural network NN1 and the second neural network NN2 do not satisfy the predetermined condition (NO in S55), the information processing apparatus 1 performs the processing after S32 again.

[0132] On the other hand, in a case where it is determined that the first neural network NN1 and the second neural network NN2 satisfy the predetermined condition (YES in S55), the information processing apparatus 1 ends the main processing of the determination model generation processing.

[Data Estimation Processing (1)]

[0133] Next, in the determination model generation processing, a first example of processing of determining whether or not contents of image data IM2 and contents of character string data ST2 that are included in the new training data DT3 match (hereinafter, also referred to as data estimation processing) will be described. FIG. 17 is a flowchart illustrating the first example of the data estimation processing.

[0134] As illustrated in FIG. 17, for example, the data reception unit 111 stands by until receiving the new training data DT3 transmitted from the operation terminal 2 (NO in S81).

[0135] Then, in the case of receiving the new training data DT3 (YES in S81), the vector generation unit 114 calculates, by using the first neural network NN1, an image vector for the image data IM2 and a character string vector for the character string data ST2 that are included in the new training data DT3 received in the processing in S81 (S82).

[0136] For example, the vector generation unit 114 inputs the new training data DT3 to the first neural network NN1. Then, the vector generation unit 114 acquires the image vector and the character string vector that are output from the first neural network NN1 as an image vector and a character string vector for the new training data DT3.

[0137] Then, for example, by calculating an inner product of the vectors calculated in the processing in S82, the similarity calculation unit 115 calculates a degree of similarity between the image data IM2 and the character string data ST2 that are included in the new training data DT3 received in the processing in S81 (S83).

[0138] Thereafter, the result output unit 118 outputs, for example, the degree of similarity calculated in the processing in S83 (S84).

[0139] For example, the result output unit 118 outputs the degree of similarity calculated in the processing in S83 to the operation terminal 2.

[0140] Note that, for example, in a case where a degree of similarity for each of a plurality of pieces of new training data DT3 is calculated in the processing in S83, the result output unit 118 may output information indicating each of the plurality of pieces of new training data DT3 in descending order of the degree of similarity calculated in the processing in S83.

[Data Estimation Processing (2)]

[0141] Next, a second example of the data estimation processing will be described. FIG. 18 is a flowchart illustrating the second example of the data estimation processing.

[0142] As illustrated in FIG. 18, for example, the data reception unit 111 stands by until receiving the new training data DT3 transmitted from the operation terminal 2 (NO in S91).

[0143] Then, in the case of receiving the new training data DT3 (YES in S91), the vector generation unit 114 calculates, by using the first neural network NN1, an image vector for the image data IM2 and a character string vector for the character string data ST2 that are included in the new training data DT3 received in the processing in S91 (S92).

[0144] Then, the matching degree calculation unit 117 calculates an element product of the vectors calculated in the processing in S92 (S93).

[0145] Moreover, the matching degree calculation unit 117 calculates, by using the second neural network NN2, a degree of matching between the image data IM2 and the character string data ST2 that are included in the new training data DT3 received in the processing in S91 (S94).

[0146] For example, the matching degree calculation unit 117 inputs the element product calculated in the processing in S94 to the second neural network NN2. Then, the matching degree calculation unit 117 acquires the degree of matching output from the second neural network NN2 as a degree of matching between the image data IM2 and the character string data ST2 that are included in the new training data DT3 received in the processing in S91.

[0147] Thereafter, the result output unit 118 outputs, for example, the degree of matching calculated in the processing in S94 (S95).

[0148] Note that, for example, in a case where a degree of matching for each of a plurality of pieces of new training data DT3 is calculated in the processing in S94, the result output unit 118 may output information indicating each of the plurality of pieces of new training data DT3 in descending order of the degree of matching calculated in the processing in S94.

[Specific Example (2) of Determination Model]

[0149] Next, a determination model MD2 in the first embodiment will be described. FIGs. 19 and 20 are diagrams illustrating processing of the determination model MD2 in the first embodiment. For example, FIGs. 19 and 20 are diagrams illustrating the processing of the determination model corresponding to the contents of FIGs. 12 to 18.

[0150] First, processing of the determination model MD2 in a learning stage will be described.

[0151] As illustrated in FIG. 19, the information processing apparatus 1 inputs, to the first neural network NN1, a pair of image data IM1 and character string data ST1 that are included in the first training data DT1 and pairs of image data IM1 and character string data ST1 that are included in the second training data DT2. For example, as illustrated in FIG. 19, the information processing apparatus 1 inputs, to the first neural network NN1, a plurality of pairs including a pair of image data IM1a and character string data ST1a that are included in the first training data DT1, a pair of image data IM1b and character string data ST1b that are included in the second training data DT2, and a pair of image data IM1c and character string data ST1c that are included in the second training data DT2.

[0152] Then, for example, in response to reception of input of the pair of image data IM1a and character string data ST1a, the first neural network NN1 calculates and outputs an image vector indicating characteristics of the image data IM1a and a character string vector indicating characteristics of the character string data ST1a. Similarly, for example, in response to reception of input of the pair of image data IM1b and character string data ST1b, the first neural network NN1 calculates and outputs an image vector indicating characteristics of the image data IM1b and a character string vector indicating characteristics of the character string data ST1b. Moreover, for example, in response to reception of input of the pair of image data IM1c and character string data ST1c, the first neural network NN1 calculates and outputs an image vector indicating characteristics of the image data IM1c and a character string vector indicating characteristics of the character string data ST1c.

[0153] Subsequently, the information processing apparatus 1 calculates, for example, an inner product of the image vector corresponding to the image data IM1a and the character string vector corresponding to the character string data ST1a as a degree of similarity between the image data IM1a and the character string data ST1a. Similarly, the information processing apparatus 1 calculates, for example, an inner product of the image vector corresponding to the image data IM1b and the character string vector corresponding to the character string data ST1b as a degree of similarity between the image data IM1b and the character string data ST1b. Moreover, the information processing apparatus 1 calculates, for example, an inner product of the image vector corresponding to the image data IM1c and the character string vector corresponding to the character string data ST1c as a degree of similarity between the image data IM1c and the character string data ST1c.

[0154] Thereafter, for example, the information processing apparatus 1 adjusts the weight of the first neural network NN1 such that a classification error calculated from the degree of similarity for each of the first training data DT1 and the second training data DT2 and information indicating the first training data DT1 (correct pair) becomes small.

[0155] Furthermore, the information processing apparatus 1 calculates, for example, an element product of the image vector corresponding to the image data IM1a and the character string vector corresponding to the character string data ST1a and an element product of the image vector corresponding to the image data IM1b and the character string vector corresponding to the character string data ST1b.

[0156] Then, the information processing apparatus 1 inputs, to the second neural network NN2, the element product corresponding to the image data IM1a and the character string data ST1a and the element product corresponding to the image data IM1b and the character string data ST1b.

[0157] Subsequently, for example, in response to reception of input of the element product corresponding to the image data IM1a and the character string data ST1a, the second neural network NN2 calculates and outputs a degree of matching between the image data IM1a and the character string data ST1a. Furthermore, for example, in response to reception of input of the element product corresponding to the image data IM1b and the character string data ST1b, the second neural network NN2 calculates and outputs a degree of matching between the image data IM1b and the character string data ST1b.

[0158] Moreover, the information processing apparatus 1 adjusts the weights of the first neural network NN1 and the second neural network NN2 such that an error between the degree of matching between the image data IM1a and the character string data ST1a and a value corresponding to the first training data DT1 (correct data) becomes small. Furthermore, the information processing apparatus 1 adjusts the weights of the first neural network NN1 and the second neural network NN2 such that an error between the degree of matching between the image data IM1b and the character string data ST1b and a value corresponding to the second training data DT2 becomes small.

[0159] Next, processing of the determination model MD2 in an inference stage will be described.

[0160] As illustrated in FIG. 20, the information processing apparatus 1 inputs, to the first neural network NN1, a pair of image data IM2 and character string data ST2 that are included in the new training data DT3.

[0161] Then, for example, in response to reception of input of the pair of image data IM2 and character string data ST2, the first neural network NN1 calculates and outputs an image vector indicating characteristics of the image data IM2 and a character string vector indicating characteristics of the character string data ST2.

[0162] Thereafter, the information processing apparatus 1 calculates, for example, an inner product of the image vector corresponding to the image data IM2 and the character string vector corresponding to the character string data ST2 as a degree of similarity between the image data IM2 and the character string data ST2.

[0163] Then, the information processing apparatus 1 outputs, for example, the calculated degree of similarity to the operation terminal 2 as a degree of similarity between the image data IM2 and the character string data ST2 that are included in the new training data DT3.

[0164] Furthermore, the information processing apparatus 1 calculates, for example, an element product of the image vector corresponding to the image data IM2 and the character string vector corresponding to the character string data ST2. Moreover, the information processing apparatus 1 inputs the calculated element product to the second neural network NN2.

[0165] Then, for example, in response to reception of input of the element product corresponding to the image data IM2 and the character string data ST2, the second neural network NN2 calculates and outputs a degree of matching between the image data IM2 and the character string data ST2.

[0166] Thereafter, the information processing apparatus 1 outputs, for example, the degree of matching output from the second neural network NN2 to the operation terminal 2 as a degree of matching between the image data IM2 and the character string data ST2 that are included in the new training data DT3.

[0167] As described above, the information processing apparatus 1 in the present embodiment generates, on the basis of the first training data DT1 in which the image data IM1 and the character string data ST1 corresponding to the image data IM1 are associated with each other, the second training data DT2 by replacing one of the image data IM1 and the character string data ST1 that are included in the first training data DT1 with another data.

[0168] Then, the information processing apparatus 1 generates, by using the first training data DT1 and the second training data DT2 as input data, a determination model that outputs information indicating which piece of training data in the first training data DT1 and the second training data DT2 is training data in which correspondence between the image data IM1 and the character string data ST1 is correct.

[0169] For example, in a case where a plurality of pairs of image data IM1 and character string data ST1 to be input to the determination model is generated, the information processing apparatus 1 in the present embodiment combines different character string data ST1 with the same image data IM1. Furthermore, in this case, the information processing apparatus 1 combines different image data IM1 with the same character string data ST1. Then, the information processing apparatus 1 trains the determination model by using the generated plurality of pairs.

[0170] With this configuration, the information processing apparatus 1 may cause the determination model to learn a detailed relationship between pieces of the training data. Therefore, the information processing apparatus 1 may generate the determination model that accurately determines correspondence between the image data IM1 and the character string data ST1.

[0171] Note that, in the examples described above, a case has been described where the first neural network NN1 that generates image vectors from the image data IM1 and the image data IM2 and the first neural network NN1 that generates character string vectors from the character string data ST1 and the character string data ST2 are the same neural network, but these may be different neural networks from each other.

[0172] With this configuration, the information processing apparatus 1 may train the first neural network NN1 that generates the image vectors by using only the image data IM1, and may also train the first neural network NN1 that generates the character string vectors by using only the character string data ST1. Therefore, in this case, the information processing apparatus 1 does not need to train the first neural network NN1 by inputting the image data IM1 and the character string data ST1 at the same time, and may efficiently generate the first neural network NN1.

Claims

1. A determination model generation program that causes at least one computer to execute a process, the process comprising:

generating, based on first training data in which image data and character string data that corresponds to the image data are associated with each other, second training data by replacing one of data included in the first training data selected from the image data and the character string data with another data; and
 generating, by using the first training data and the second training data as input data, a determination model that outputs information that indicates which training data selected from the first training data and the second training data is training data in which correspondence between the image data and the character string data is correct.

2. The determination model generation program according to claim 1, wherein
 generating the second training data includes generating, for each of a plurality of the first training data, a plurality of the second training data by replacing one of data included in each of the plurality of the first training data selected from the image data and the character string data with another data.

3. The determination model generation program according to claim 2, wherein
 the generating the second training data includes

generating, for each of a first part of the plurality of the first training data, the plurality of the second training data by replacing the image data included in each of the first part of the plurality of the first training data with another data, and
 generating, for each of a second part of the plurality of the first training data, the plurality of the second training data by replacing the character string data included in each of the second part of the plurality of the first training data with another data.

4. The determination model generation program according to claim 1, wherein
 the generating the determination model includes

calculating, for each of the first training data and the second training data, a degree of similarity between the image data and the character string data included in each of the first training data and the second training data, and
 using the degree and first information that indicates which training data selected in the first training data and the second training data is the first training data.

5. The determination model generation program according to claim 4, wherein
 the calculating includes calculating, for each of the first training data and the second training data, an inner product of the image data and the character string data included in each of the first training data and the second training data as the degree.

6. The determination model generation program according to claim 4, wherein the generating the determination model includes
 using the degree and the first information and second information that indicates whether or not the correspondence in each of the first training data and the second training data is correct.

7. An information processing apparatus comprising:

a data generation unit that generates, based on first training data in which image data and character string data that corresponds to the image data are associated with each other, second training data by replacing one of data included in the first training data selected from the image data and the character string data with another data; and
 a model training unit that generates, by using the first training data and the second training data as input data, a determination model that outputs information that indicates which training data selected from the first training data and the second training data is training data in which correspondence between the image data and the character string data is correct.

8. The information processing apparatus according to claim 7, wherein the data generation unit that further generates the second training data includes generating, for each of a plurality of the first training data, a plurality of the second training data by replacing one of data included in each of the plurality of the first training data selected from the image data and the character string data with another data.

9. The information processing apparatus according to claim 8, wherein the data generation unit that further

generates, for each of a first part of the plurality of the first training data, the plurality of the second training data by replacing the image data included in each of the first part of the plurality of the first training data with another data, and

generates, for each of a second part of the plurality of the first training data, the plurality of the second training data by replacing the character string data included in each of the second part of the plurality of the first training data with another data.

10. The information processing apparatus according to claim 7, wherein the model training unit that further

calculates, for each of the first training data and the second training data, a degree of similarity between the image data and the character string data included in each of the first training data and the second training data, and uses the degree and first information that indicates which training data selected in the first training data and the second training data is the first training data.

11. A determination model generation method for a computer to execute a process comprising:

generating, based on first training data in which image data and character string data that corresponds to the image data are associated with each other, second training data by replacing one of data included in the first training data selected from the image data and the character string data with another data; and

generating, by using the first training data and the second training data as input data, a determination model that outputs information that indicates which training data selected from the first training data and the second training data is training data in which correspondence between the image data and the character string data is correct.

12. The determination model generation method according to claim 13, wherein

generating the second training data includes generating, for each of a plurality of the first training data, a plurality of the second training data by replacing one of data included in each of the plurality of the first training data selected from the image data and the character string data with another data.

13. The determination model generation method according to claim 14, wherein

the generating the second training data includes

generating, for each of a first part of the plurality of the first training data, the plurality of the second training data by replacing the image data included in each of the first part of the plurality of the first training data with another data, and

generating, for each of a second part of the plurality of the first training data, the plurality of the second training data by replacing the character string data included in each of the second part of the plurality of the first training data with another data.

14. The determination model generation method according to claim 13, wherein

the generating the determination model includes

calculating, for each of the first training data and the second training data, a degree of similarity between the image data and the character string data included in each of the first training data and the second training data, and using the degree and first information that indicates which training data selected in the first training data and the second training data is the first training data.

FIG. 1

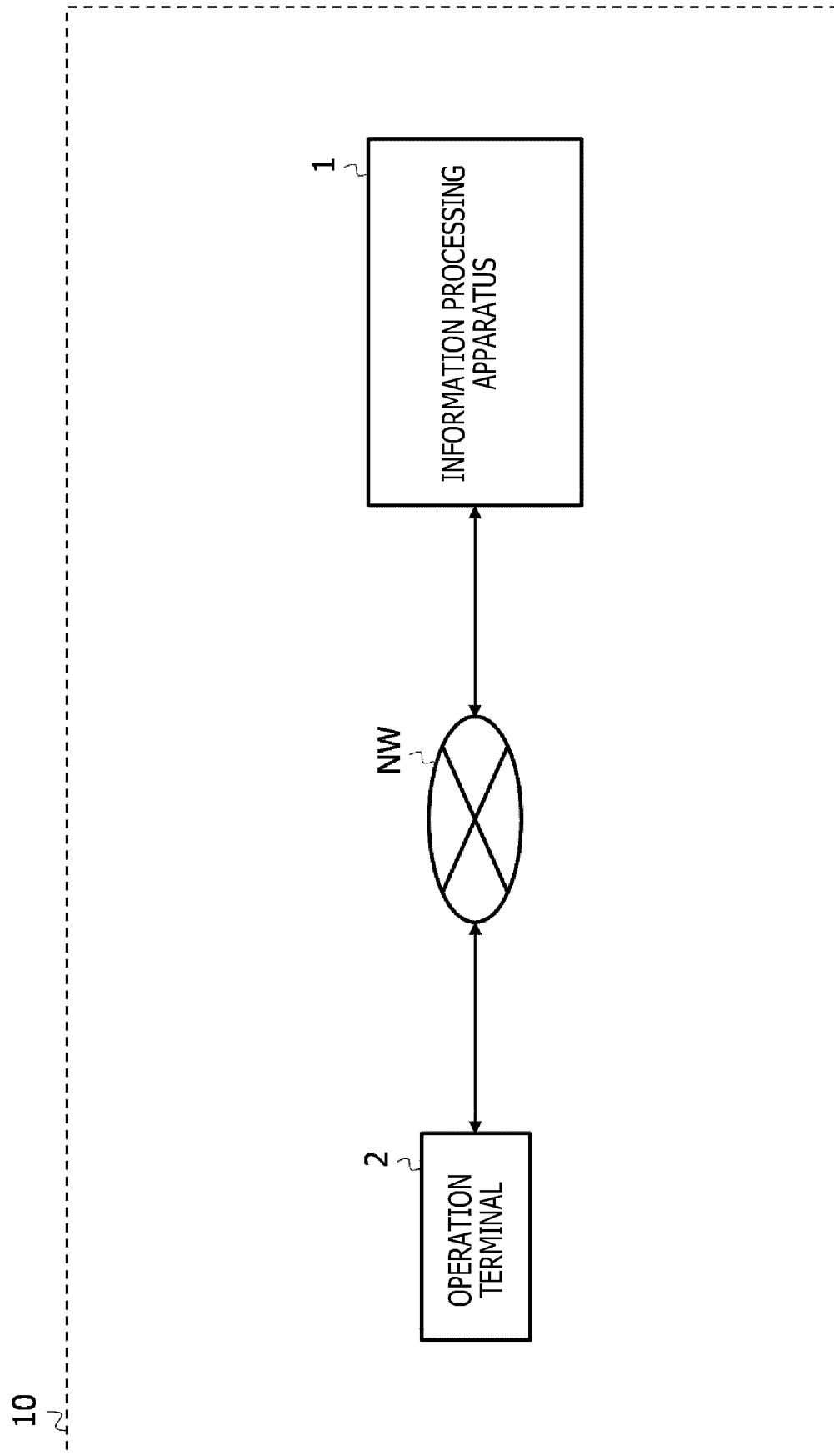


FIG. 2

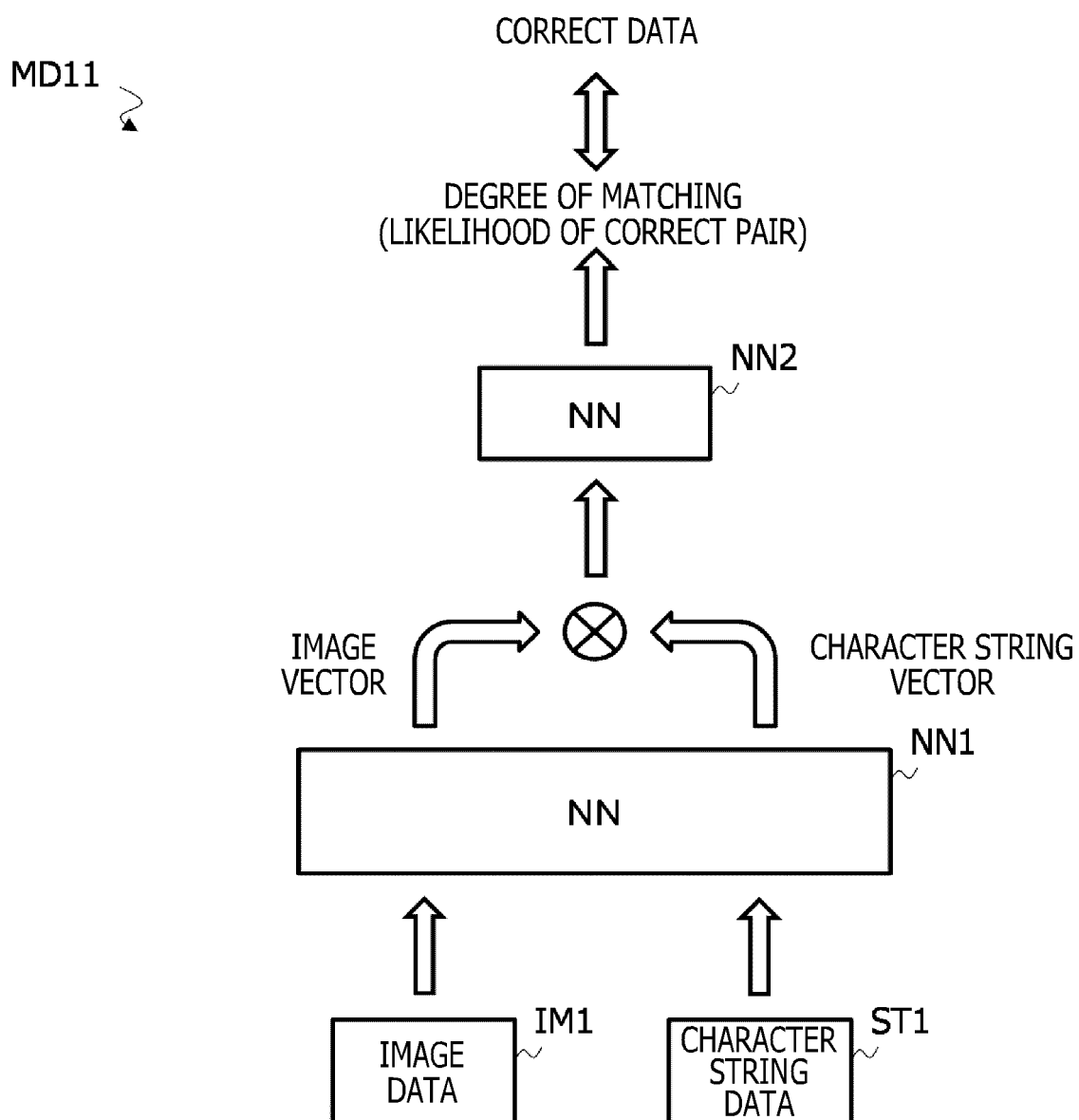


FIG. 3

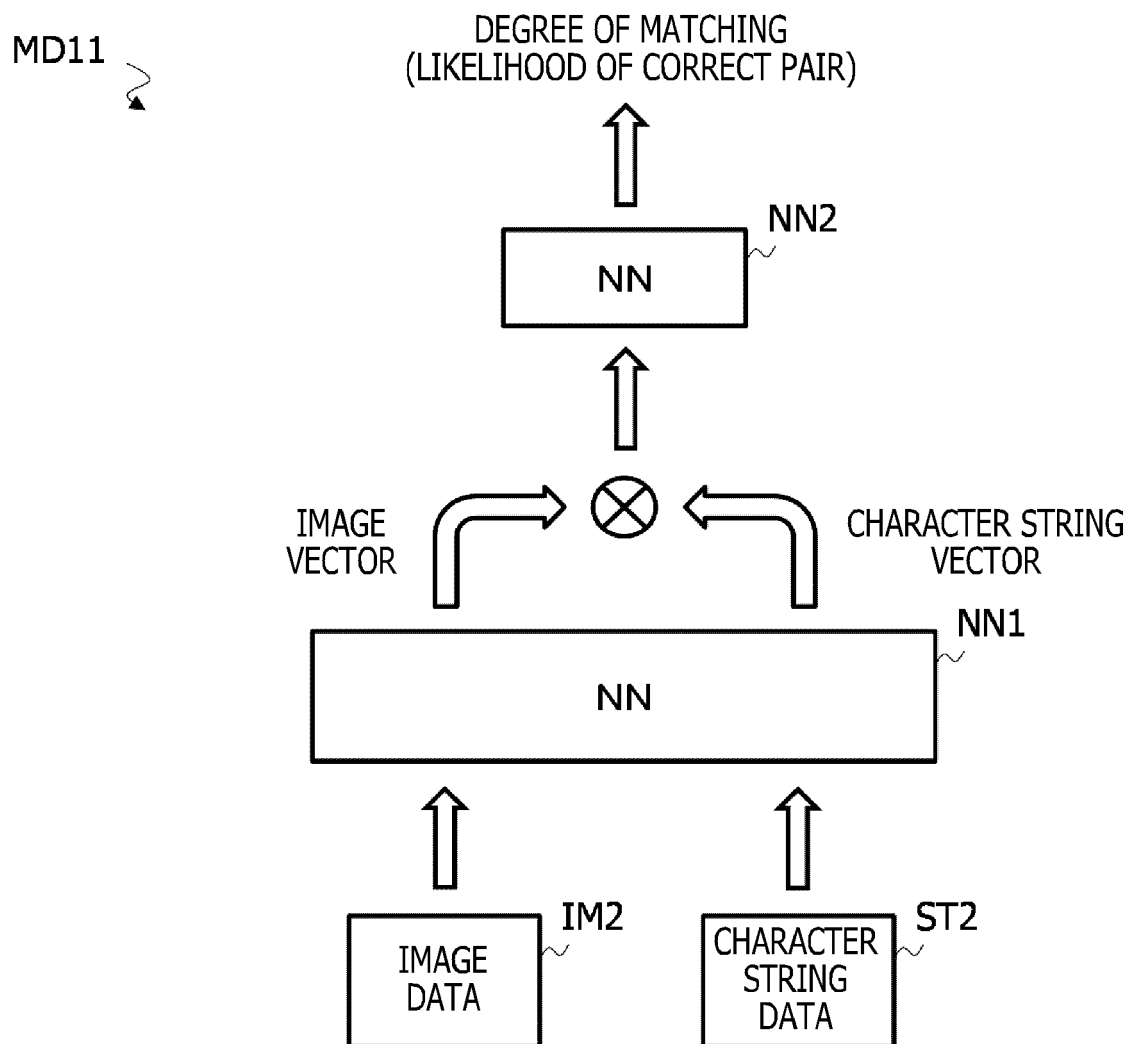


FIG. 4

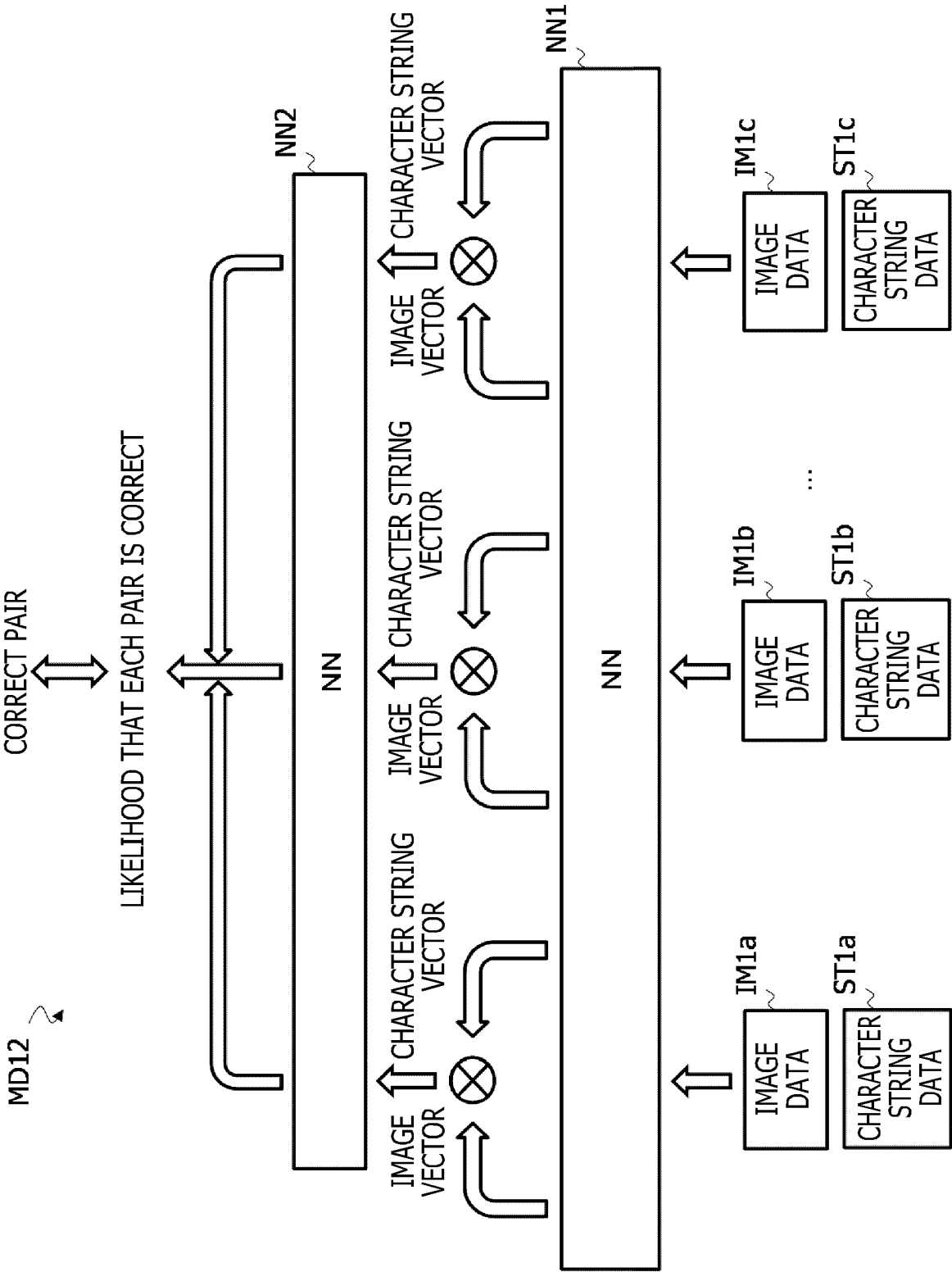


FIG. 5

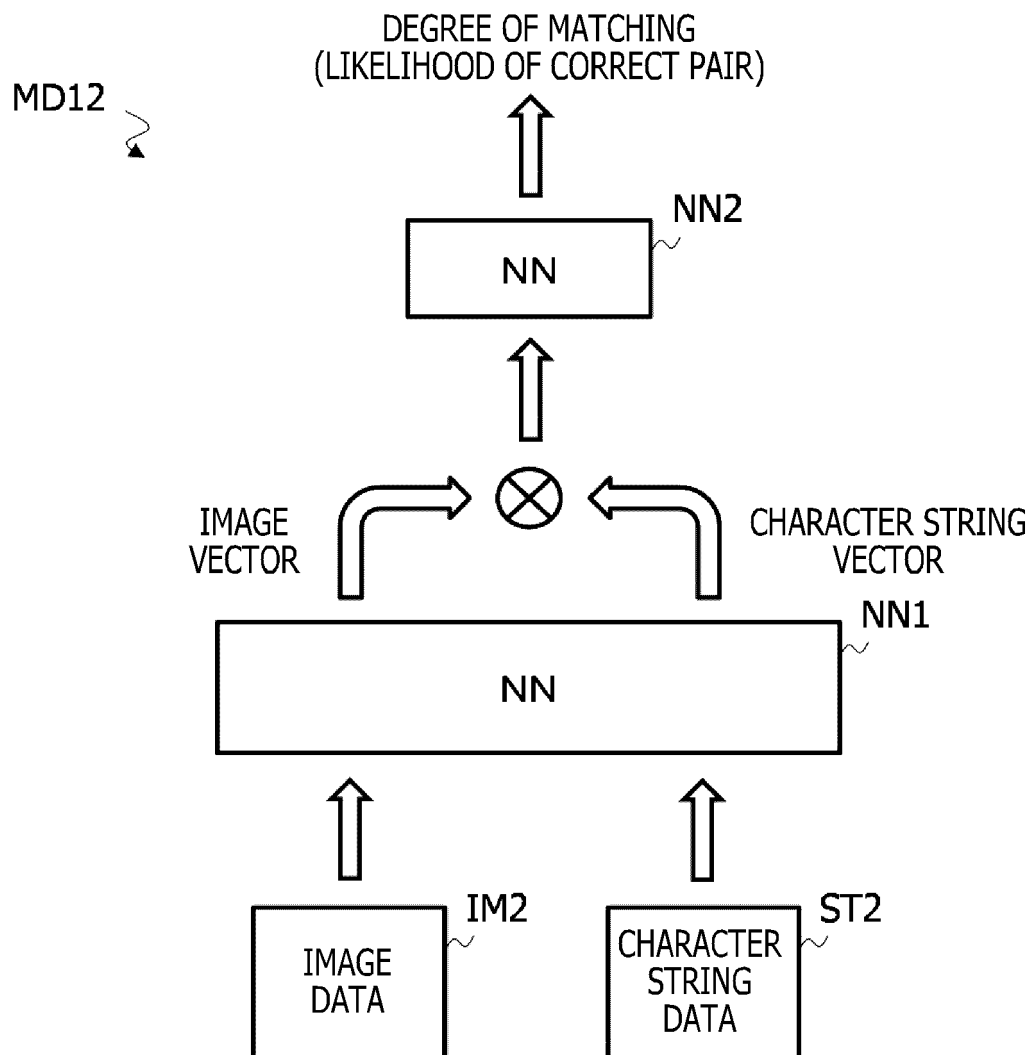


FIG. 6

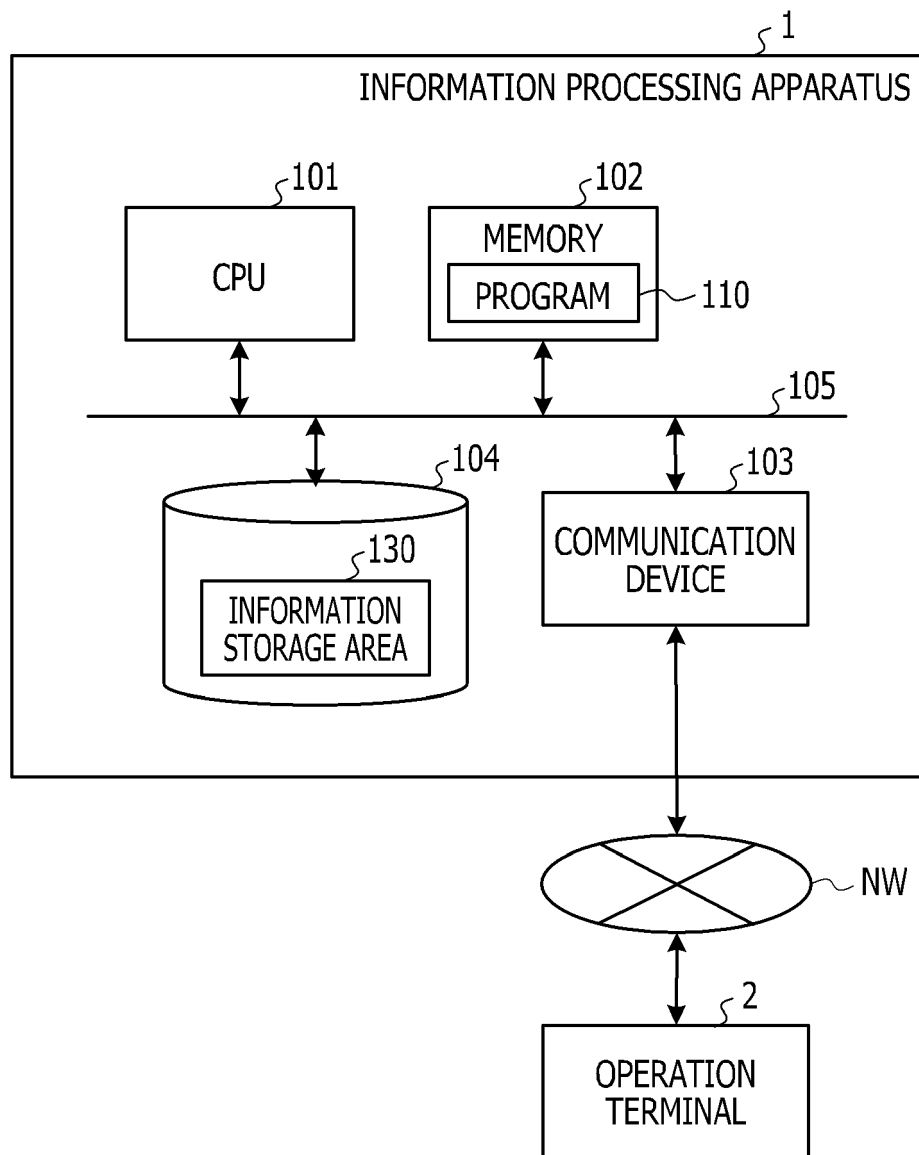


FIG. 7

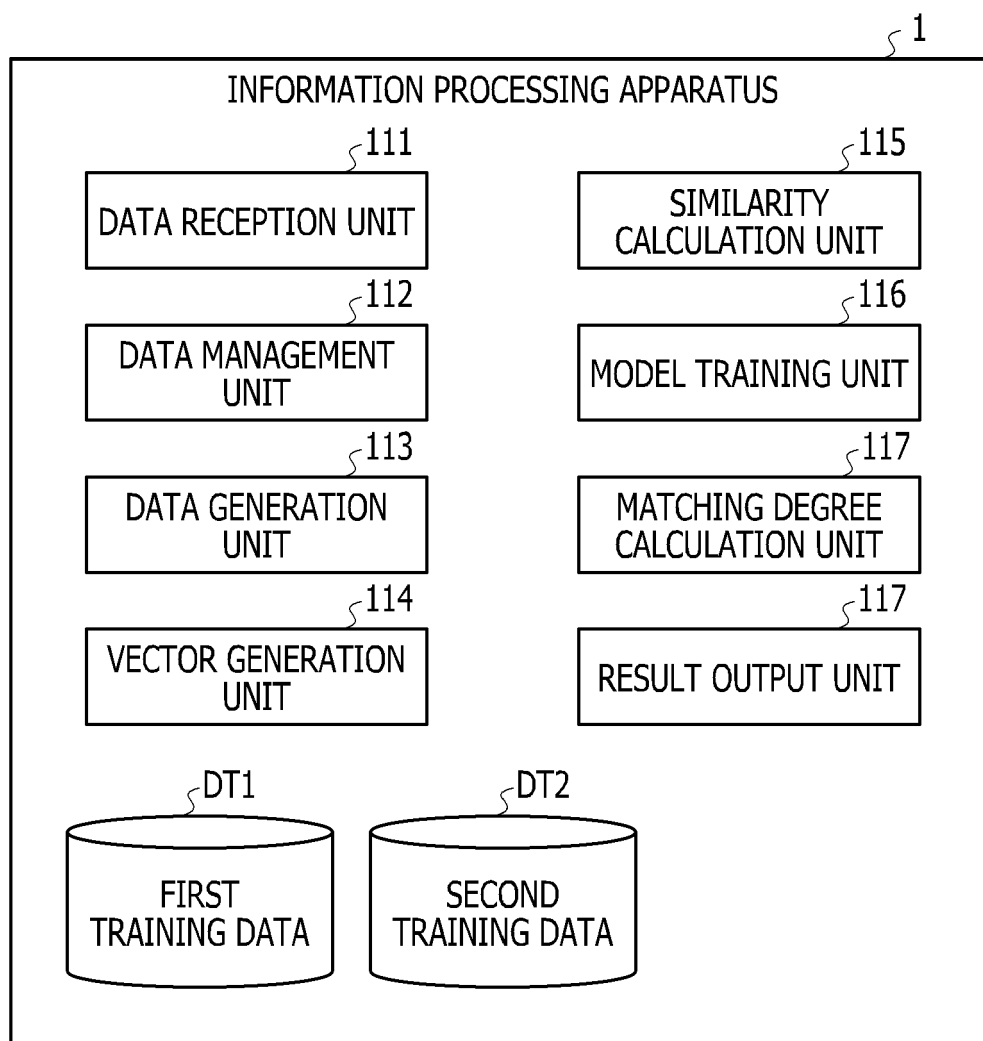


FIG. 8

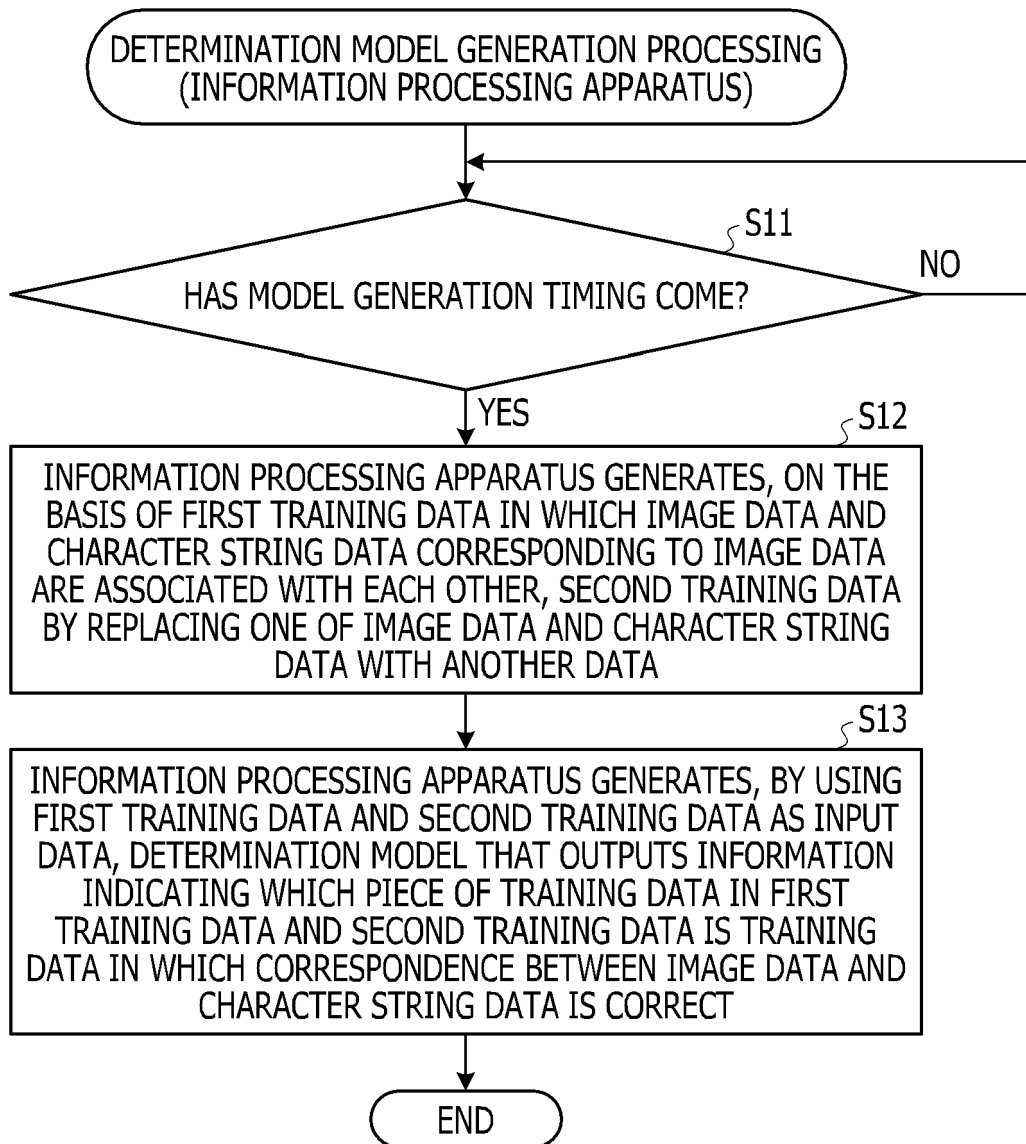


FIG. 9

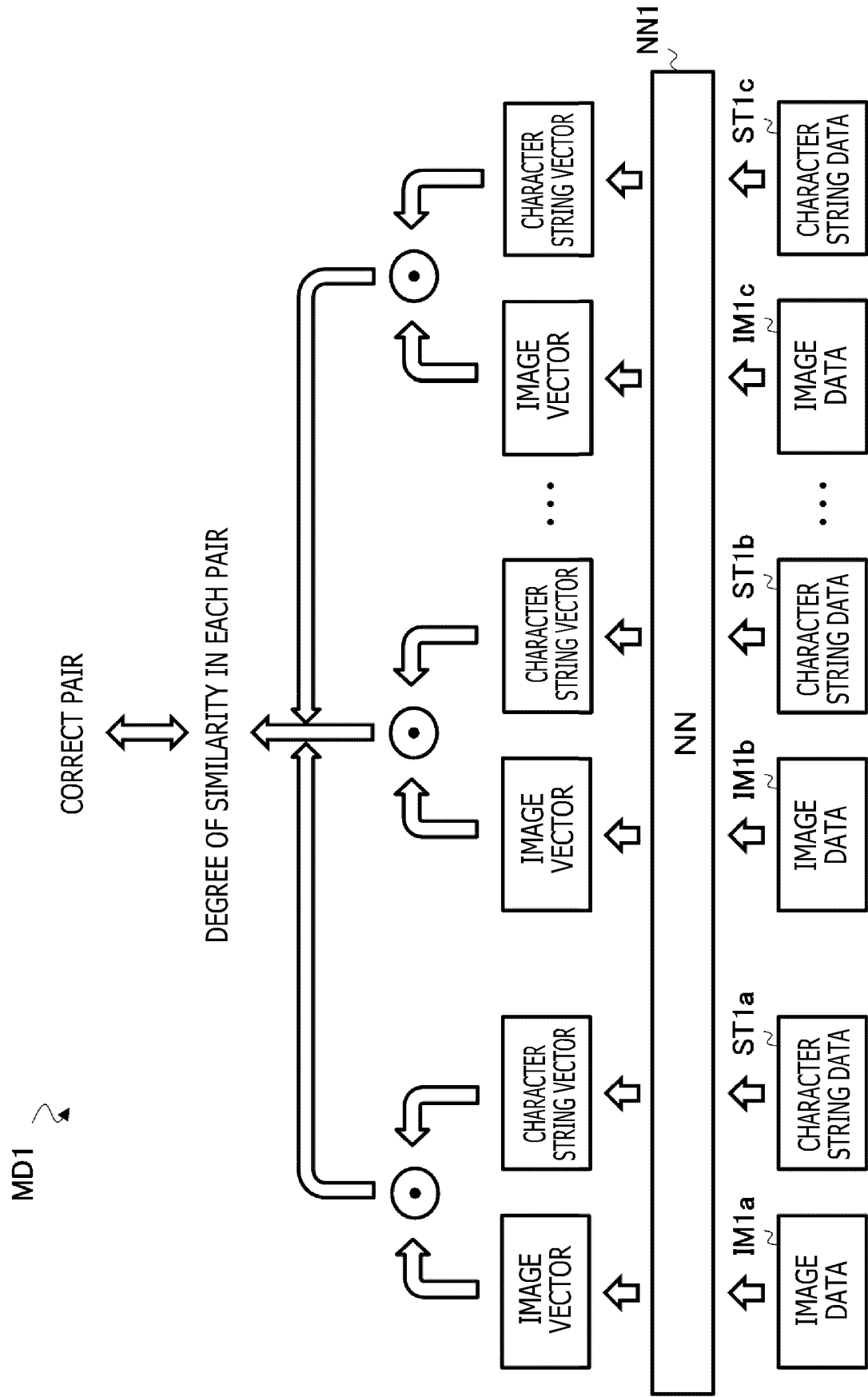


FIG. 10

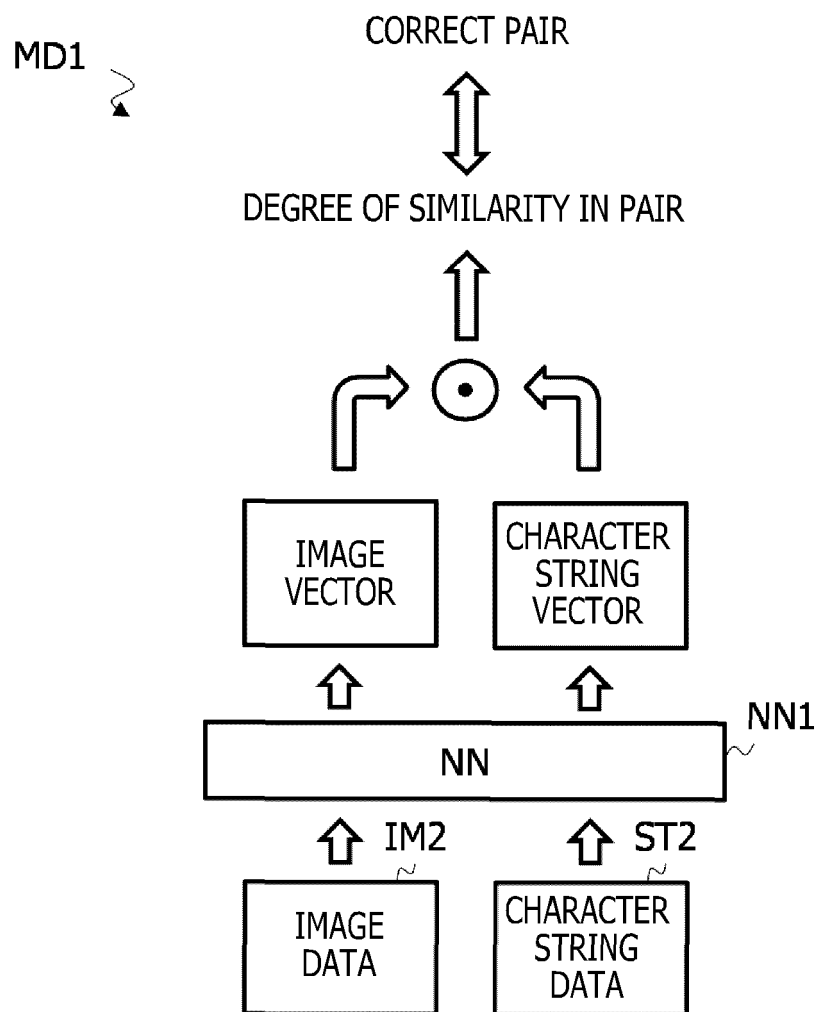


FIG. 11

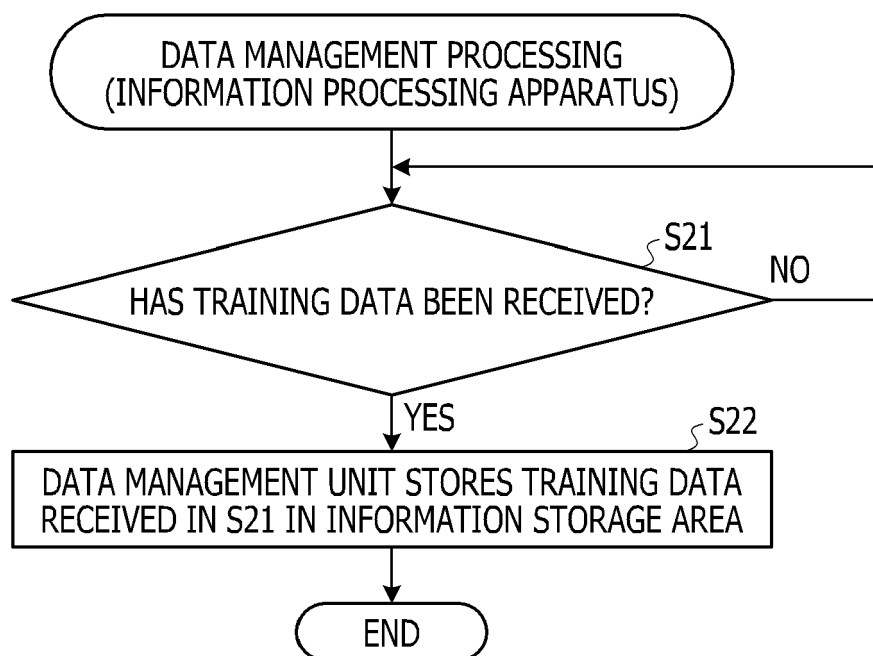


FIG. 12

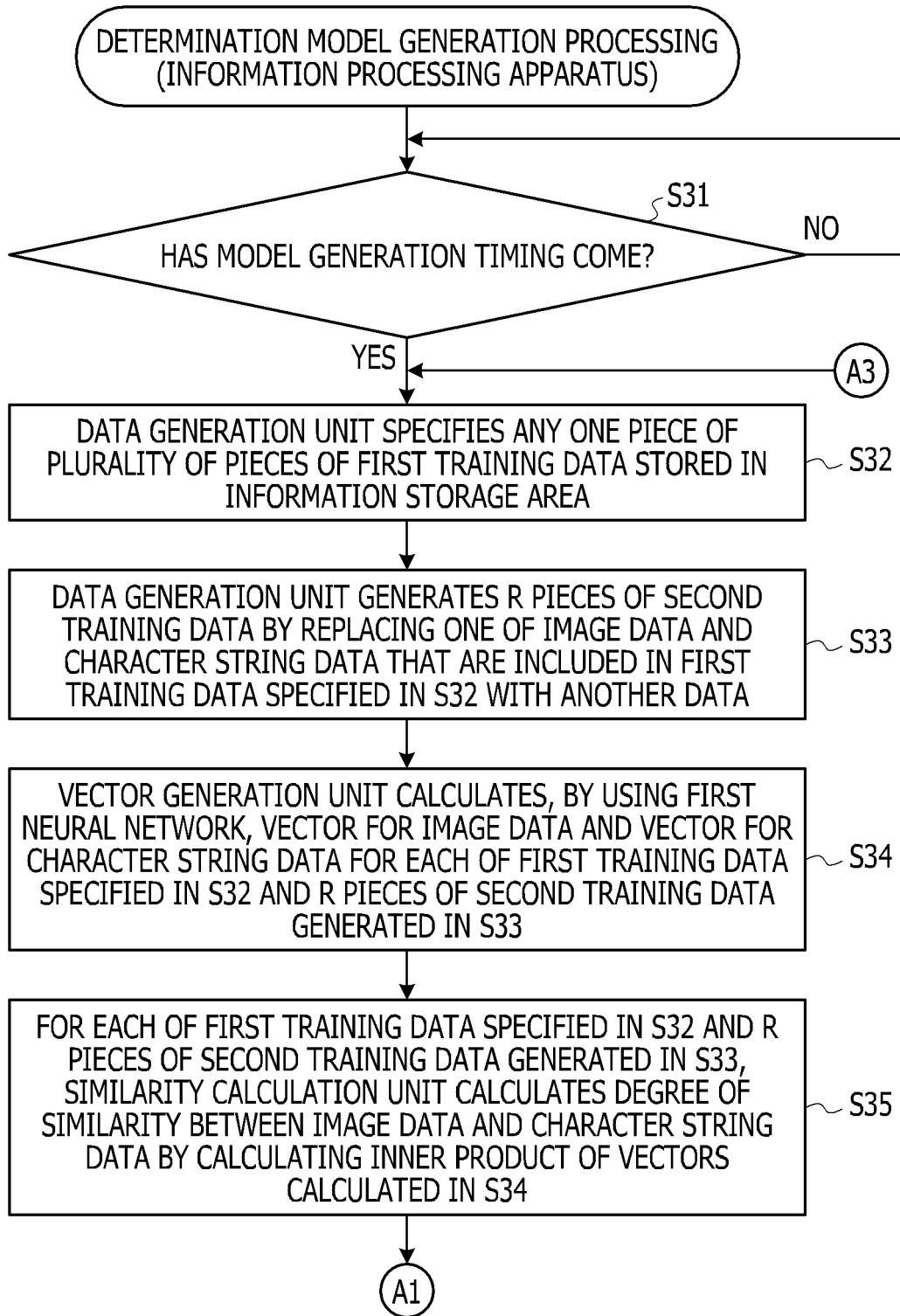


FIG. 13

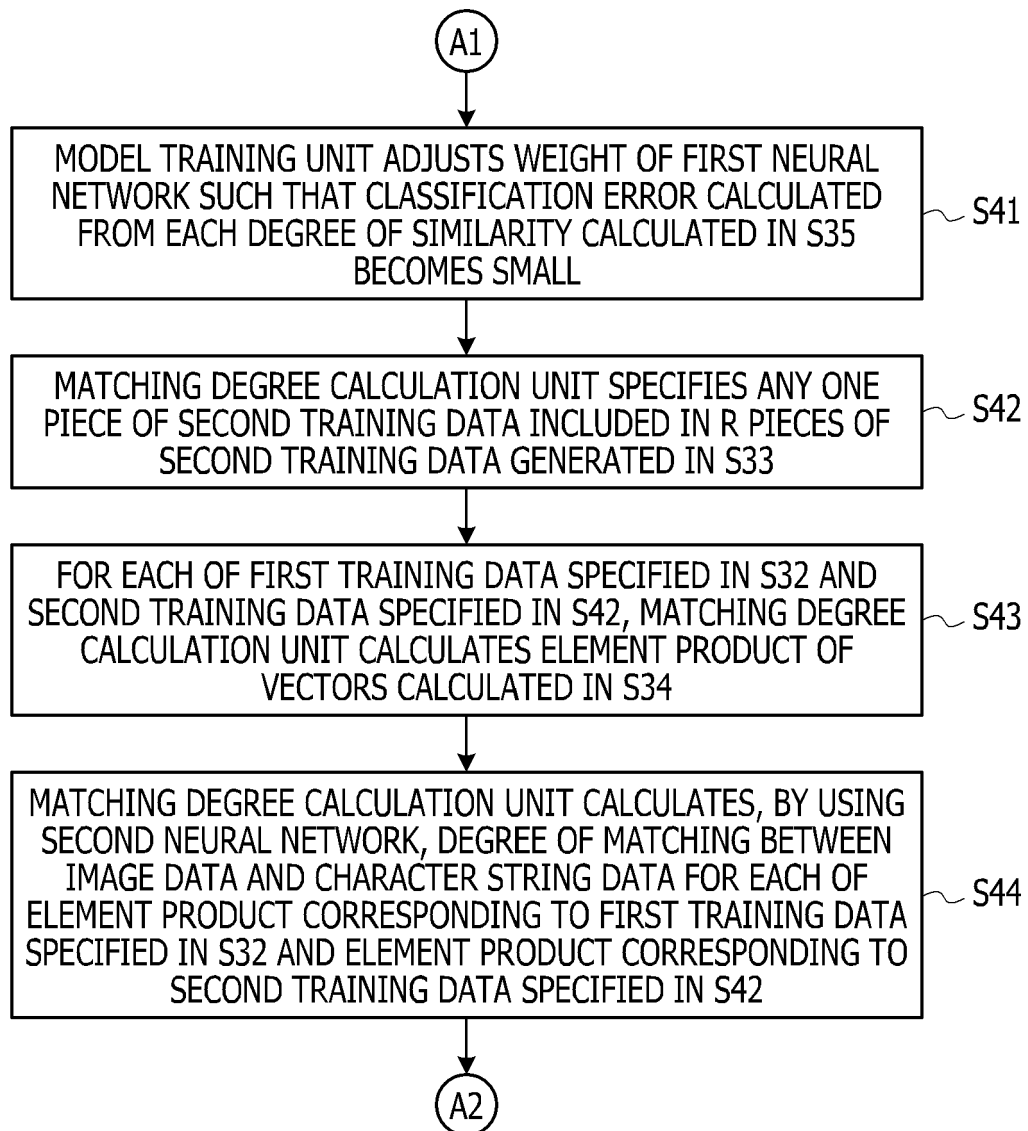


FIG. 14

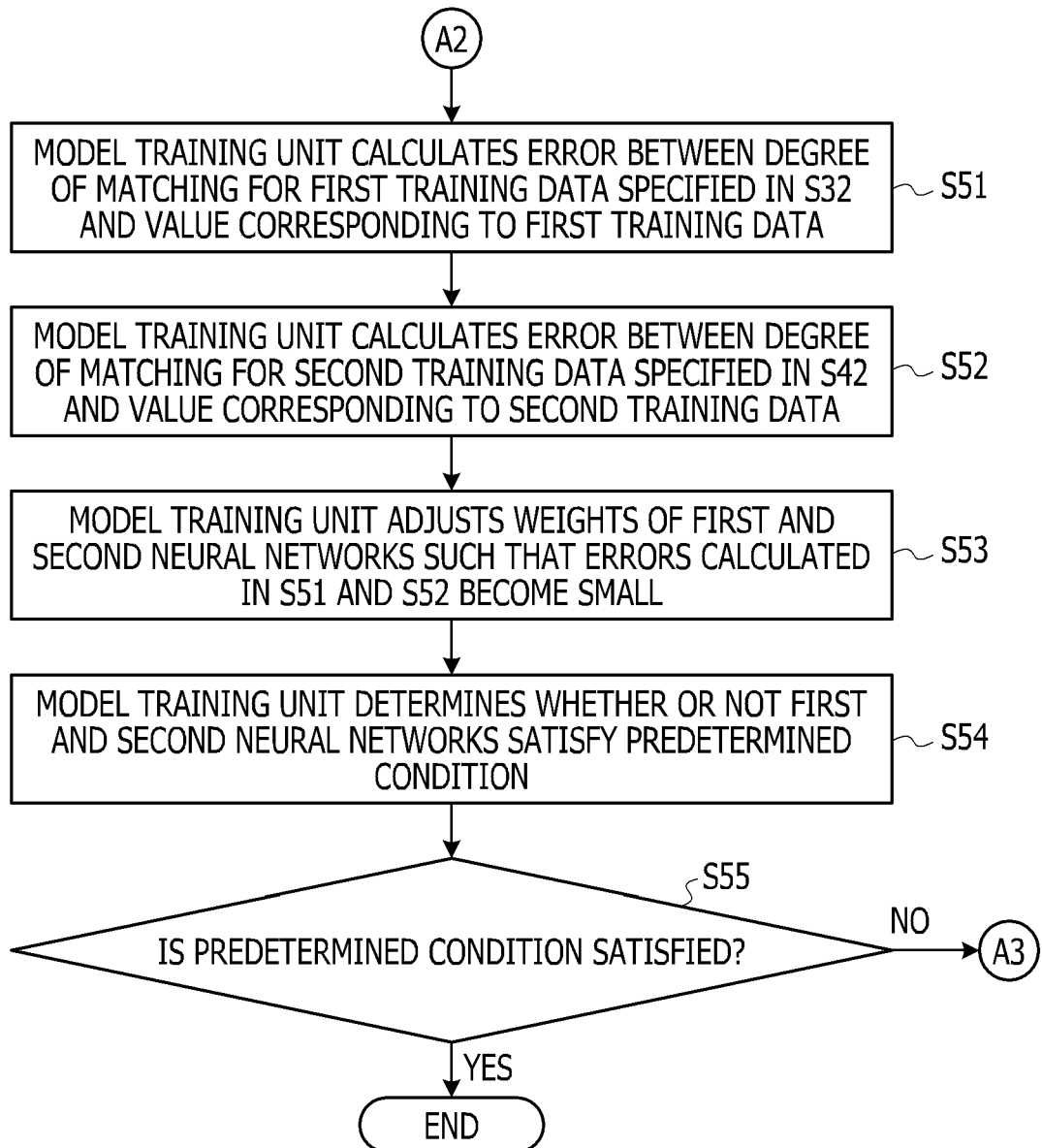


FIG. 15

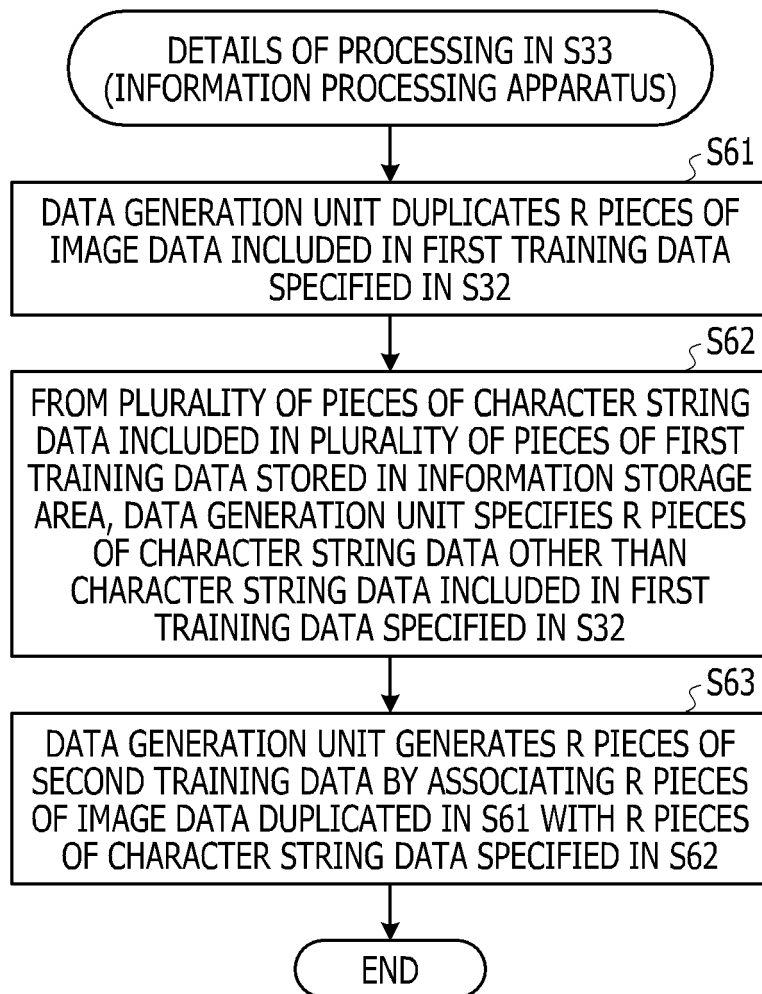


FIG. 16

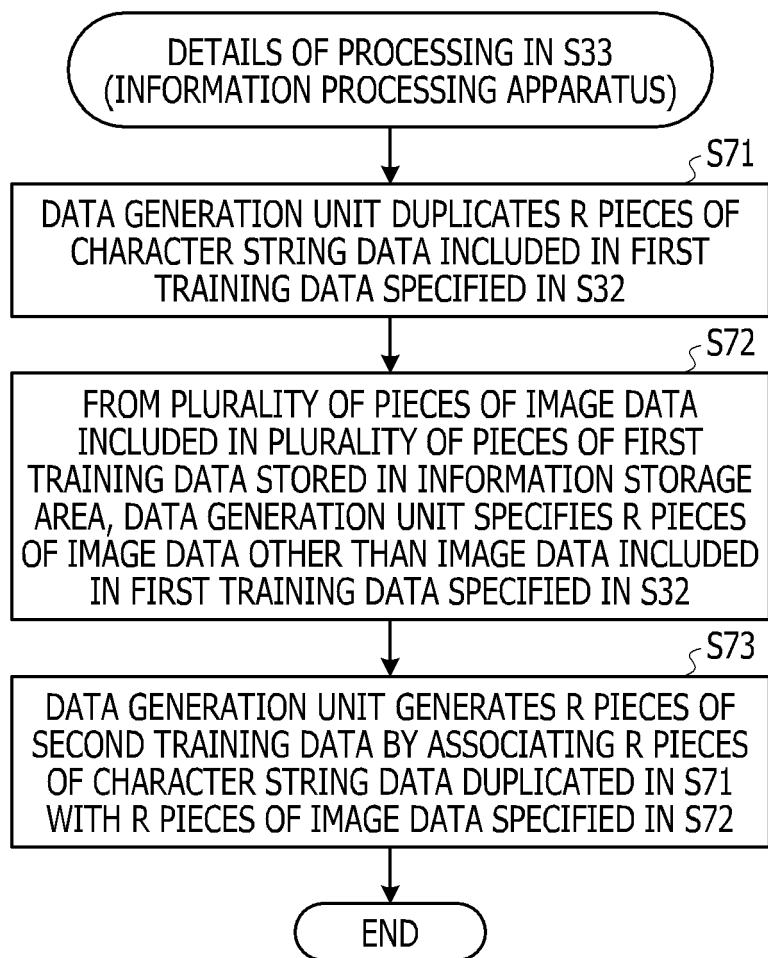


FIG. 17

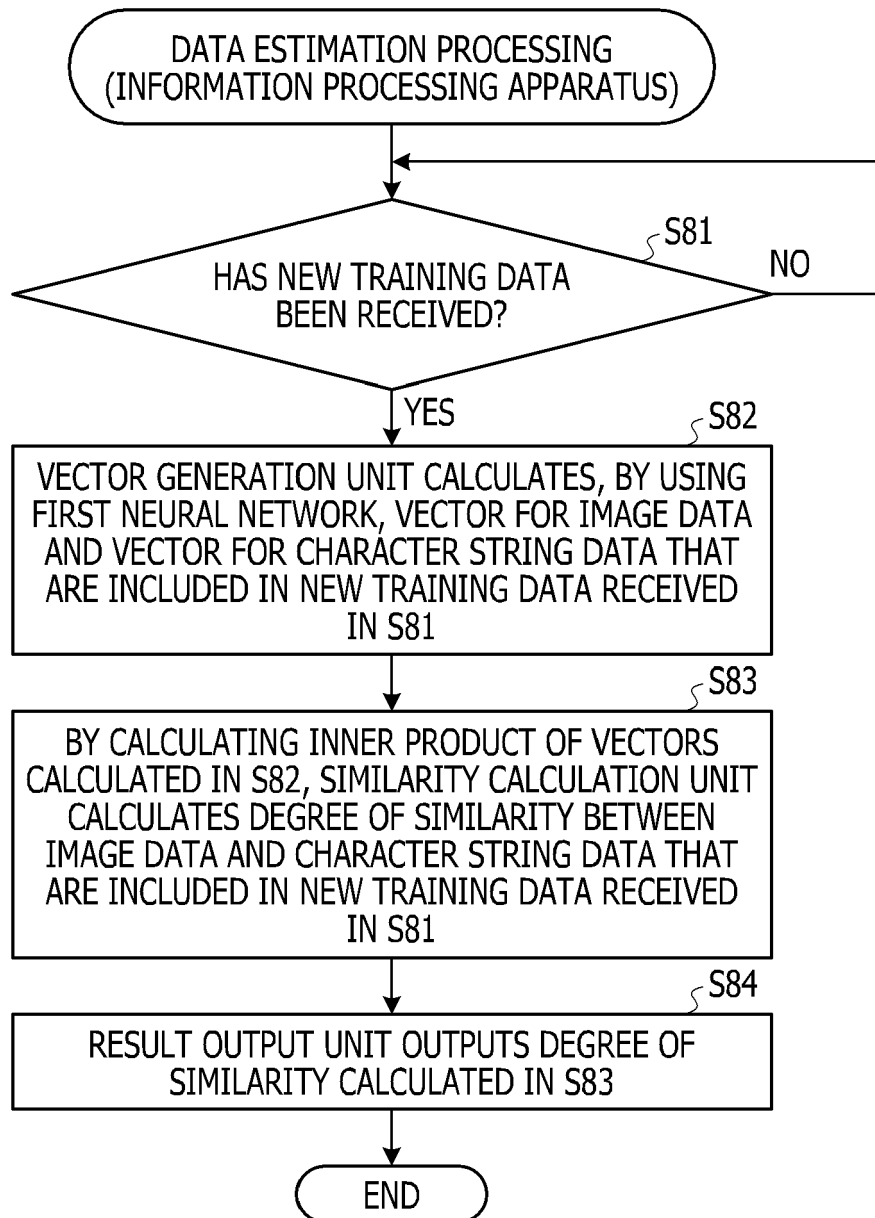


FIG. 18

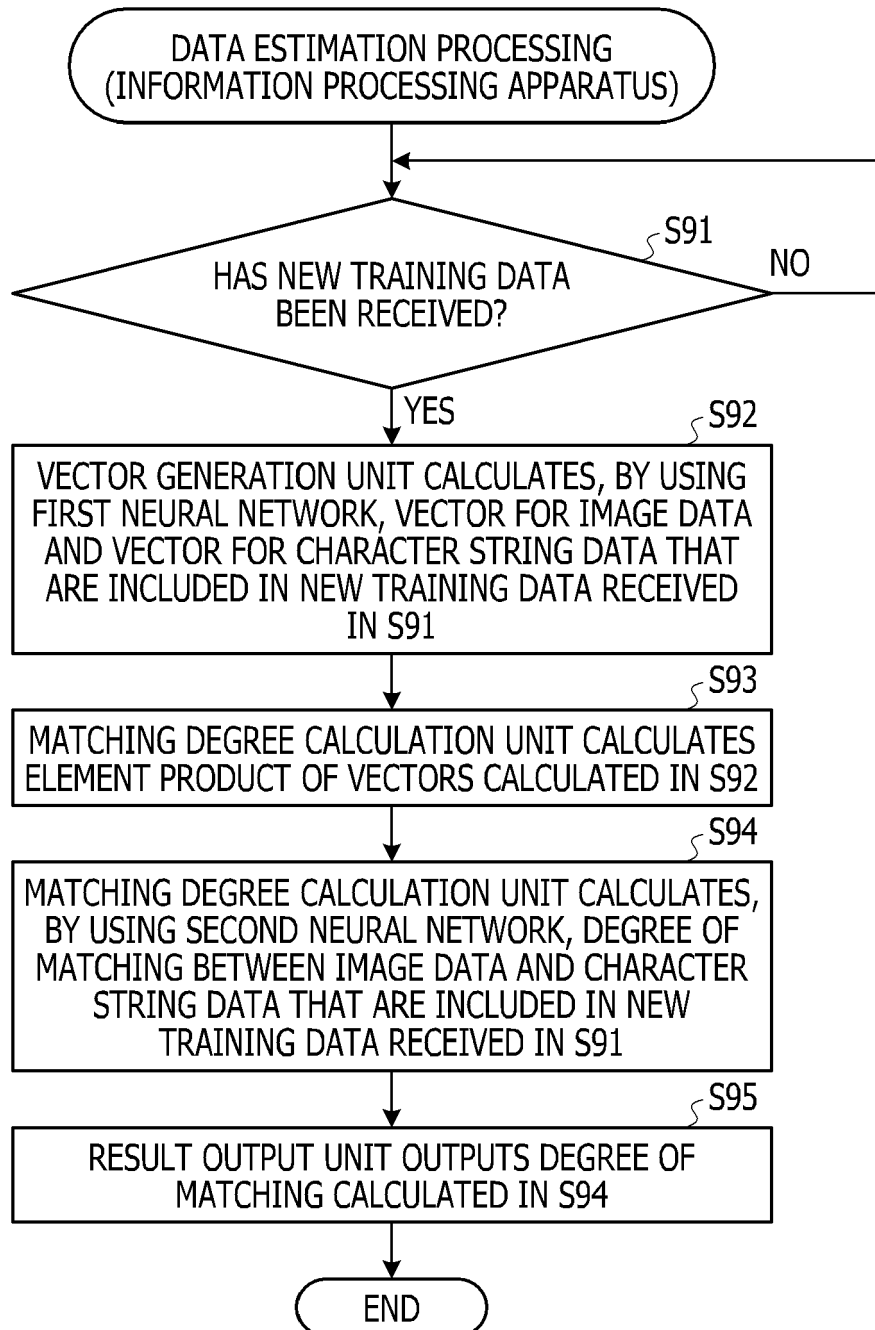


FIG. 19

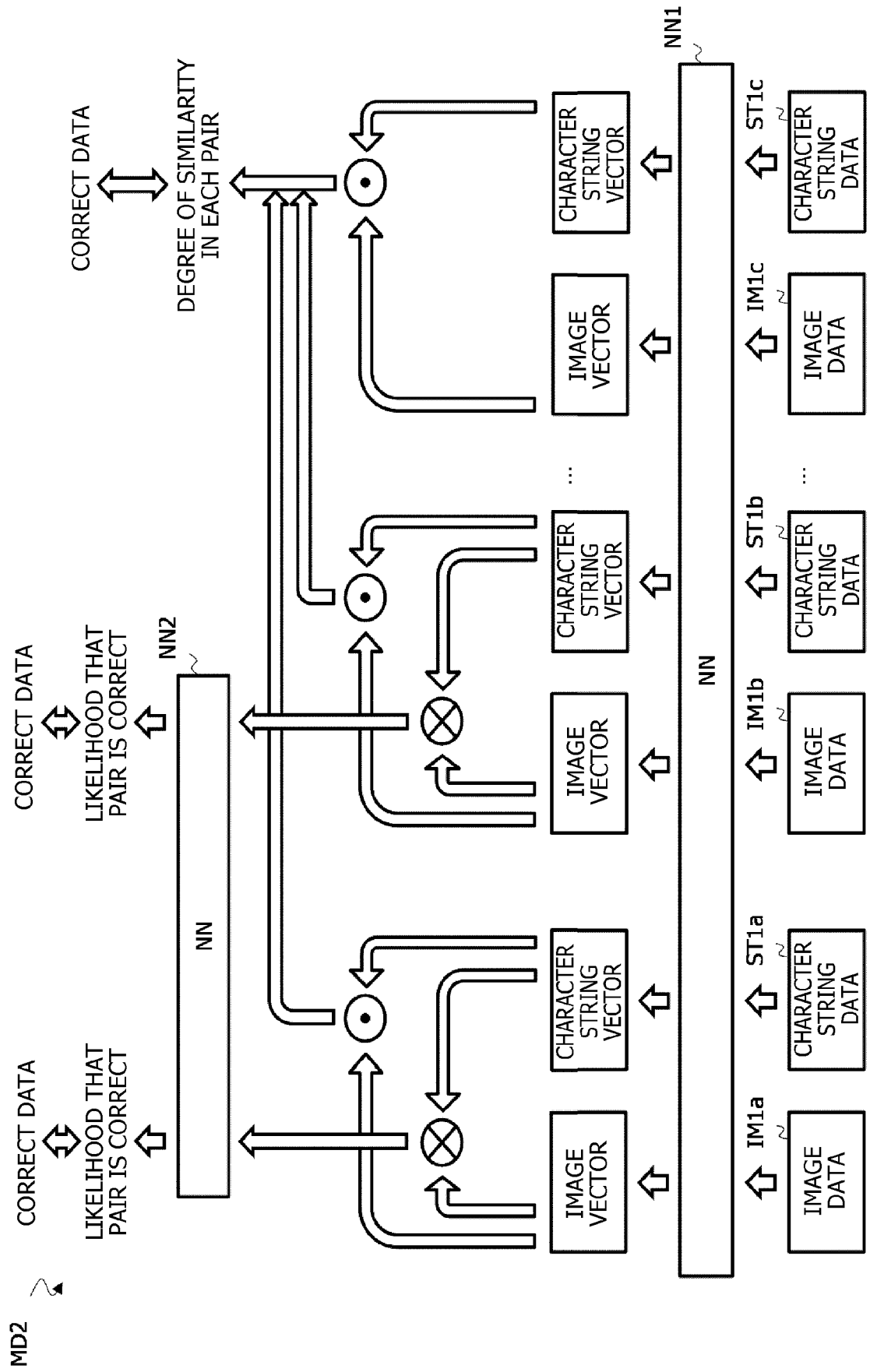
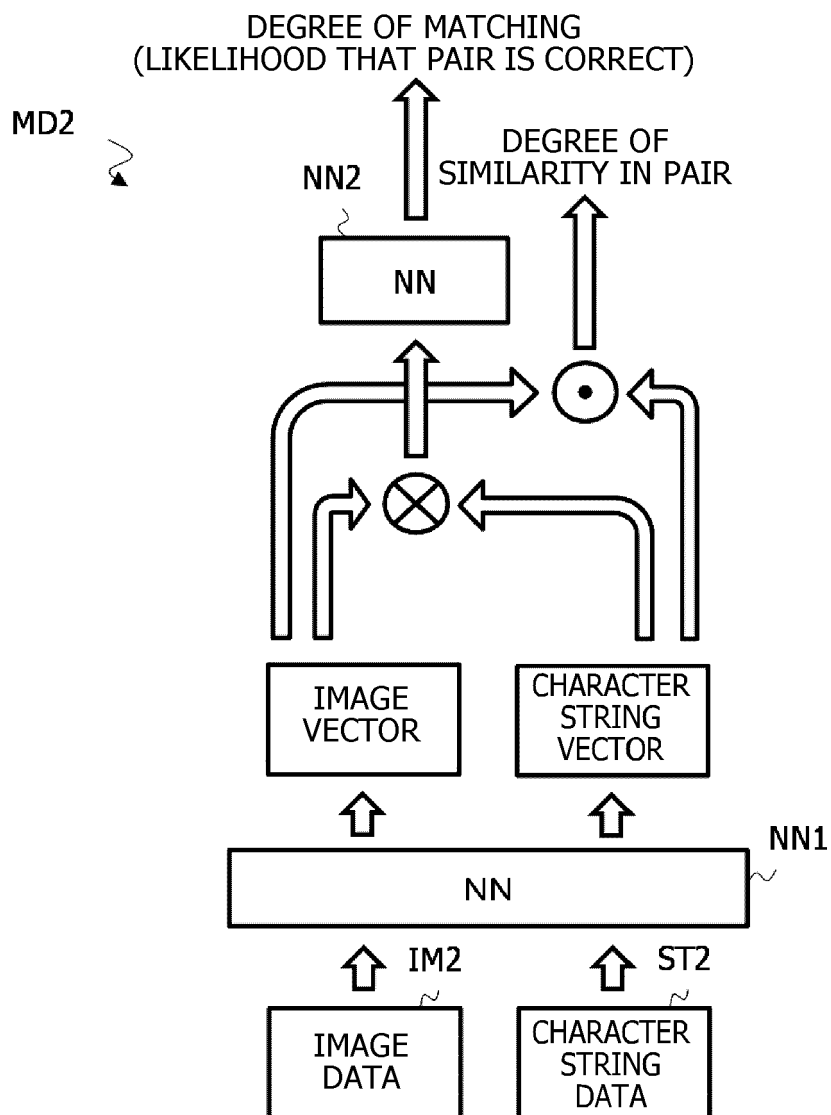


FIG. 20





EUROPEAN SEARCH REPORT

Application Number

EP 21 20 3124

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	<p>TIANLANG CHEN ET AL: "Adaptive Offline Quintuplet Loss for Image-Text Matching", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 22 July 2020 (2020-07-22), XP081706212, * abstract; figures 1, 2 *</p> <p>* section 3.3 *</p> <p>* section 4.2 *</p> <p>* section 4.3 *</p> <p>* section 7.1 *</p> <p>* page 3, line 7 - line 16 *</p> <p>* page 7, line 8 - line 13 *</p> <p>* page 10, line 8 - line 23 *</p> <p>* equations 3, 4, 5 *</p> <p>-----</p>	1-14	<p>INV.</p> <p>G06N3/08</p> <p>G06N3/04</p>
X	<p>LI KUNPENG ET AL: "Visual Semantic Reasoning for Image-Text Matching", 2019 IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV), IEEE, 27 October 2019 (2019-10-27), pages 4653-4661, XP033723868, DOI: 10.1109/ICCV.2019.00475 [retrieved on 2020-02-24]</p> <p>* abstract *</p> <p>* section 1, lines 4-6 *</p> <p>* section 3.4 *</p> <p>* section 4 *</p> <p>* equation 8 *</p> <p>-----</p> <p style="text-align: center;">-/--</p>	1-14	<p>TECHNICAL FIELDS SEARCHED (IPC)</p> <p>G06N</p>
The present search report has been drawn up for all claims			
Place of search		Date of completion of the search	Examiner
The Hague		19 April 2022	Aoun, Marc
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone</p> <p>Y : particularly relevant if combined with another document of the same category</p> <p>A : technological background</p> <p>O : non-written disclosure</p> <p>P : intermediate document</p> <p>T : theory or principle underlying the invention</p> <p>E : earlier patent document, but published on, or after the filing date</p> <p>D : document cited in the application</p> <p>L : document cited for other reasons</p> <p>& : member of the same patent family, corresponding document</p>			



EUROPEAN SEARCH REPORT

Application Number

EP 21 20 3124

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X,D	<p>JIASEN LU ET AL: "ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 6 August 2019 (2019-08-06), XP081456681, * abstract; figures 3b, 4 *</p> <p>* section 3.1 *</p> <p>* section 3.2 *</p> <p>* page 4, line 36 - line 43 *</p> <p>-----</p>	1-14	
The present search report has been drawn up for all claims			TECHNICAL FIELDS SEARCHED (IPC)
Place of search The Hague			Date of completion of the search 19 April 2022
			Examiner Aoun, Marc
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

 2
 EPO FORM 1503 03.82 (P04C01)